2008

Modeling Training Effects on Task Performance Using a Human Performance Taxonomy

Douglas P. Meador
Wright State University

MODELING TRAINING EFFECTS ON TASK PERFORMANCE USING A HUMAN PERFORMANCE TAXONOMY

A dissertation submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

By

DOUGLAS PAUL MEADOR
B.S., Miami University, 1985
M.S.E., Wright State University, 2005

2008
Wright State University

WRIGHT STATE UNIVERSITY

SCHOOL OF GRADUATE STUDIES

November 20, 2008

I HEREBY RECOMMEND THAT THE DISSERATION PREPARED UNDER MY SUPERVISION BY Douglas Paul Meador ENTITLED Modeling Training Effects on Task Performance Using a Human Performance Taxonomy BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Doctor of Philosophy.

Raymond R. Hill, Ph.D.
Dissertation Director

Ramana Grandhi, Ph.D.
Director, Ph.D. in Engineering Program

Joseph F. Thomas, Jr., Ph.D.
Dean, School of Graduate Studies

Committee on Final Examination

Raymond R. Hill, Ph.D.

Misty A. Blue, Ph.D.

Gerald P. Chubb, Ph.D.

Frank W. Ciarallo, Ph.D.

Christopher R. Hale, Ph.D.

Yan Liu, Ph.D.

ABSTRACT

Meador, Douglas Paul. Ph.D., Department of Biomedical, Industrial and Human Factors Engineering, Ph.D. in Engineering Program, Wright State University, 2008. Modeling training effects on task performance using a human performance taxonomy.

There is little argument that modern military systems are very complex, both in terms of the resources in time and money to develop them and the infrastructure that is required to produce trained operators. To properly execute human systems integration during the acquisition process, systems built to train operators must be developed that optimize training. Consequently, the training system community would benefit from simulation models that provide the ability to make accurate predictions of training processes, and allow the decision maker to specify an optimum combination of operator performance after training and the cost of that training. The goal of this research is the construction of a model of human learning using time to complete a task as a performance measure. While past research has explored the nature of functions to characterize human learning, this study will examine processes used to build a model that considers task performance as a function of training methods used to instruct a task, the nature of the task being taught, and the ability of the human to retain skill over a specified period of nonuse. An empirical study was performed to collect data from individuals completing tasks typically performed by sensor operators assigned to military unmanned aircraft systems. The tasks performed covered a range of activities that require varying combinations of human perceptual, cognitive and motor skills. The data were fitted to a set of models that were used to predict the performance outcome of a task similar in type to those used to build the model. Results are reported and recommendations for future research are offered.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

While performing this research and preparing this document, I took pause on several occasions to review other dissertations in hopes of learning something new about organization, process and writing style. I have taken particular interest in others' acknowledgement pages, because this activity has proven to me that no one can complete an effort of this magnitude without the gracious contributions of others.

I would like to thank Dr. Dee Andrews and Patricia Fitzgerald for their gracious approval in allowing me to use data collected during the IMPRINT Training Algorithm Enhancements (ITAE) project performed by Science Applications International Corporation (SAIC). Their generosity provided me with a huge volume of data without which this research could not have been accomplished.

I would also like to thank my coworkers at SAIC who participated in the ITAE study. SAIC ITAE program manager Jeffrey Doyal gave me the latitude to create an experimental design, collect data, and analyze it in a manner that makes this research my unique contribution to process modeling. Sue Knight and Dena Trammell at the SAIC Huntsville location built very professional computer-based training materials that formed the backbone of one of the independent variables in this research. Michael Sargent and Jennifer Devilbiss built custom-made data collection software that made collection and storage of data for analysis a breeze, an absolute must for an experiment of this magnitude.

Next, I would like to express my gratitude to my dissertation committee:

> Dr. Ray Hill for his direction of this dissertation and the tremendous effort expended to help me get this dissertation on paper.
>
> Drs. Frank Ciarallo, Misty Blue and Yan Liu for serving on my committee and providing an academic challenge that made this effort a true learning experience.
>
> Dr. Jerry Chubb for traveling on several occasions from The Ohio State University and providing unique insights to my understanding of the state of the art of training and modeling.
>
> Dr. Chris Hale for taking time from his workload at SAIC to participate and providing a readily available sounding board for ideas and questions.

Last, and certainly not least, I would like to thank my wife, Tricia, and children Stephen, Katie and Lauren, for their patience during the long hours it took to complete this work.

1. INTRODUCTION

Military organizations employ training operations unsurpassed in their complexity and expense. This assertion is supported by the statement made by Orlansky et al., (1985):

> *Military training makes a large and continuing demand on resources allocated to the military services. For example, the time spent by students in individual training at designated schools plus that of the instructors needed at those schools accounts for about one-fourth of all the man-years (both military and civilian) available to the Department of Defense. About 20 percent of all military personnel are in schools at all times, either as students or instructors. Most (about 76 percent) of this effort simply provides initial training to new personnel entering military service for the first time.*

The enormity of military training necessitates a logical approach to determining the appropriate allocation of training assets and the proper selection of training methods. Rising development costs of modern military weapon systems – and the expanded timelines over which that development is accomplished – have forced the Department of Defense (DOD) to develop system acquisition policies that ensure an efficient system engineering process. The Defense Acquisition Guidebook (2006) requires acquisition managers ensure proper human systems integration (HSI) by "optimizing total system performance and minimizing the cost of ownership through a 'total system approach' to acquisition management." The human element of this total system approach is considered in each of seven domains of HSI:

manpower, personnel, training, human factors, safety and occupational health, personnel survivability, and habitability (DOD Instruction 5000.2, 2003).

Regarding the training domain, DOD documents emphasize that the total systems approach includes not only the actual weapon system, but the personnel who operate and maintain it and the tools (training and training devices) they are provided to learn how to do so. To support this directive, those professionals involved in the design of training systems – industrial and human factors engineers, operations research analysts, training designers, etc. – need effective tools to model and analyze training systems while the system they support is still on the drawing table as well as during its development and operational lifecycle.

One such modeling tool is IMPRINT (Improved Performance Research Integration Tool) Pro developed by the Army Research Laboratory's (ARL) Human Research and Engineering Directorate. IMPRINT Pro is a "dynamic, stochastic discrete event network modeling tool designed to help assess the interaction of soldier and system performance throughout the system lifecycle" (Allendar, 2000). The software provides the capability to decompose a broad task into a network of smaller tasks with task resolution set at the discretion of the modeler (how "specific" the modeler wants the task represented). See Figure 1 for an example IMPRINT Pro task network model.

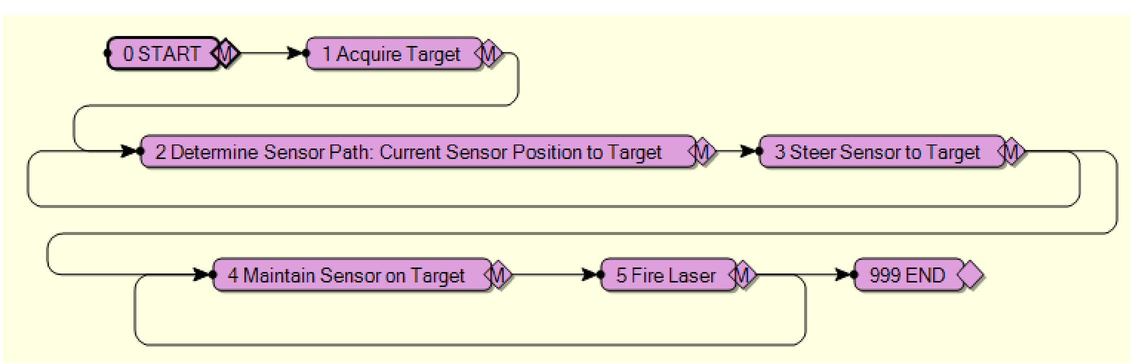


Figure 1. Notional laser designation task network model in IMPRINT Pro.

The model illustrated in Figure 1 describes a general task, tracking a target and illuminating it with a laser designator, decomposed into several smaller tasks. Each of these smaller tasks can be characterized in terms of the time and accuracy in which they are accomplished (accuracy defined in two ways – either the ability to perform a task to a specific criterion, or the probability of success of completing the task). Accuracy during the simulation can be estimated as an expression; time can be coded as an expression or called from a distribution from a graphic user interface (GUI) (see Figure 2). As the simulation runs, values produced for time and accuracy are evaluated against time and accuracy requirements and results are provided for use in system evaluation.

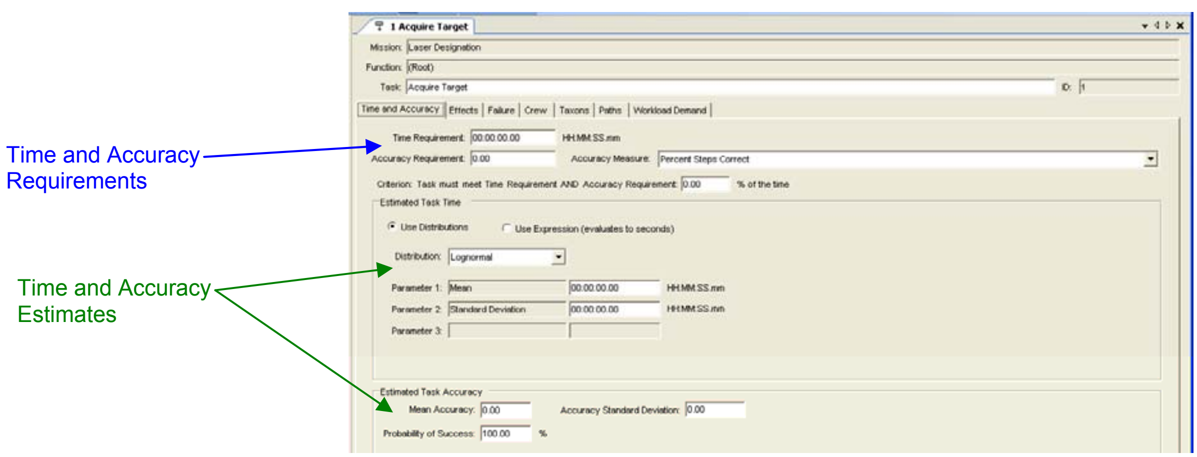


Figure 2. Time and accuracy requirements and estimates in IMPRINT Pro.

Embedded in IMPRINT Pro are functions that modify performance in terms of time to complete the task and the accuracy in which the task is accomplished. These performance modifier functions (PMF) are simple models that increase task time and decrease task accuracy as a result of two variables. The first variable is the nature of the stressor that is expected to affect performance. IMPRINT Pro's current stressor library includes cold, heat, noise, lack of sleep, and when protective gear for nuclear, biological and chemical defense is worn. The second variable is the nature of the task being accomplished and is characterized by a taxonomy of human performance. Nine “taxons” are defined that fall into four categories and map to the human capacity to receive information, process it, and then act upon it. Specifically, these categories are perceptual, cognitive, motor and communication. Table 1 lists the nine taxons with brief descriptions.

Table 1. IMPRINT Pro taxons (adapted from Alion, 2007).

Taxon	Taxon Category	Definition
Visual recognition and discrimination	Perception	Requires using the eyes to identify or separate targets or objects
Numerical analysis	Cognitive	Requires performing arithmetical or mathematical calculations
Information processing and problem solving	Cognitive	Requires processing information mentally and reaching a conclusion
Fine motor – discrete	Motor	Requires performing a set of distinct actions in a predetermined sequence mainly involving movement of the hands, arms, or feet with little physical effort
Fine motor – continuous	Motor	Requires uninterrupted performance of an action needed to keep a system on a desired path or in a specific location
Gross motor – heavy	Motor	Requires expending extensive physical effort or exertion to perform an action
Gross motor – light	Motor	Requires moving the entire body to perform an action without expending extensive physical effort
Reading and writing	Communication	Requires either reading text or numbers that are written or writing text or numbers
Oral	Communication	Requires either talking or listening to another person

In an IMPRINT Pro model task network, individual tasks may be characterized by a combination of up to three taxons. Each taxon selected is weighted according to a proportion that the modeler assumes the taxon impacts task accomplishment. For example, consider the laser designation task depicted in Figure 1 – the five composite tasks that comprise the overall task could be weighted as shown in Table 2.

Table 2. Notional taxon and taxon weight breakdown for a sample task.

Task	Taxon (CATEGORY)	Time (seconds)	Taxon Weight
Acquire Target	Visual Recognition and Discrimination (PERCEPTUAL)	5	0.20
Determine Sensor Path	Information processing and problem solving (COGNITIVE)	2	0.08
Steer Sensor to Target	Fine Motor – Continuous (MOTOR)	10	0.72
Maintain Sensor on Target	Fine Motor – Continuous (MOTOR)	5	
Fire Laser	Fine Motor – Continuous (MOTOR)	3	
TOTAL		25	1.00

While IMPRINT Pro provides a robust capability to model task performance under a variety of stressors, there is an underlying assumption that must be made about the human operator being modeled – the operator is assumed proficient in the task being accomplished. There are no PMFs available that permit the modeler to evaluate changes in performance during the learning process.

The lack of models that characterize human performance during training (derived from empirical studies) limit the effectiveness of simulation and modeling tools. A study funded by the Air Force Research Laboratory's (AFRL) Warfighter Training Division sought to quantify models of training as a function of training strategy, a time interval between training and task performance, and a range of the IMPRINT Pro taxon set for tasks performed by Predator UAV (unmanned air vehicle) sensor operators (SO) (Fitzgerald, et al., 2008). Models were built that modified time and accuracy during repeated attempts at a task during initial task training (Meador, et al., 2008).

The models built during that study are similar in construction to those PMFs already installed in IMPRINT Pro. The models, though they modify performance according to the variables just mentioned (training type, time interval and taxon makeup), are deterministic and will modify performance by exactly the same amount regardless of the number of iterations a simulation is run. A more robust model of performance during training should support a stochastic simulation that gives the modeler/analyst a set of data that more accurately represents a population of trainees.

Because there are systematic relationships between task performance and repeated attempts at a task during the learning process, human learning patterns can be characterized mathematically. This "learning curve" is a relationship between performance and practice that suggests improvements will be large during the early stages of training, then decrease in magnitude with practice until performance stabilizes at a level of proficiency (Chapanis, 1996). Learning curves are useful for several reasons, such as comparing difficulty between tasks, setting standards for personnel selection, and as predictors of task success training criteria (Swezey & Llaneras, 1997). Mathematical characterizations of learning curves provide a set of parameters that define the curve, each of which provides insight to the learning process.

This dissertation extends the research reported by Fitzgerald, et al. (2008) and Meador, et al. (2008) in two areas. First, it is hypothesized that empirical data of individuals' performance while learning will produce learning curves that can be reduced to key parameters that will form the basis for predictions of the rate of skill acquisition, amount of skill decay during periods of nonuse of the newly learned skill, and rate of skill re-acquisition. Second, the models are constructed in a way that permits their application across a broad range of tasks. In doing so, this research makes contributions in two areas that have not been adequately discussed together – the process of human learning and modeling and simulation techniques. The dissertation addresses the current state of the art of modeling in the training area and provides a methodology for developing robust models of training effectiveness.

The remainder of this document presents a background of modeling and simulation in the acquisition of training systems, a systems view of human performance, and characteristics of human learning. A research framework for building stochastic PMFs is described, followed by data collection and analysis methodologies. Results of the data analysis are discussed in the context of development and validation of the models. Finally, the findings of this research are summarized and conclusions offered.

2. BACKGROUND

The Department of Defense is perhaps the world's largest developer of new systems (Moroney & Bittner, 1995) and has set distinct policies regarding the use of modeling and simulation in the acquisition of weapon systems. Early in acquisition strategy planning, program managers are required to determine how modeling and simulation will be used during the design, build and test phases (DAU, 2006). The foci of modeling efforts are on specific measures of effectiveness for the system under study. DOD policy further directs managers of acquisition programs to ensure proper human systems integration (HSI) in seven areas: manpower, personnel, training, human factors, safety and occupational health, personnel survivability and habitability.

2.1. Modeling and Simulation in the Acquisition of Training Systems

Previous generations of weapon systems followed consistent processes during development; systems were designed, tested, and initial fielding began before significant effort was expended on determining how to train system operators and what training devices were needed. To correct problems with training issues being addressed too late in the acquisition process, DOD policy mandates several HSI activities that include the development of training devices and procedures (DOD 5000.2, 2003). Specifically, acquisition managers must "develop options for individual, collective and joint training for operators" and to "base training decisions on training effectiveness evaluations" that result in better-trained operators, improved skill retention, and reduced training costs (DOD 5000.2, 2003).

To support this mandate, acquisition managers are forced to make choices in the development of training programs. Rouse (1990) suggests these choices depend on the training objectives of interest, the characteristics of the population to be trained, and the resources required to conduct the training. Ames and Sondhi (1988) describe a technology classification model that attempts to correlate the fidelity of a spectrum of training devices with knowledge transfer (Figure 3).

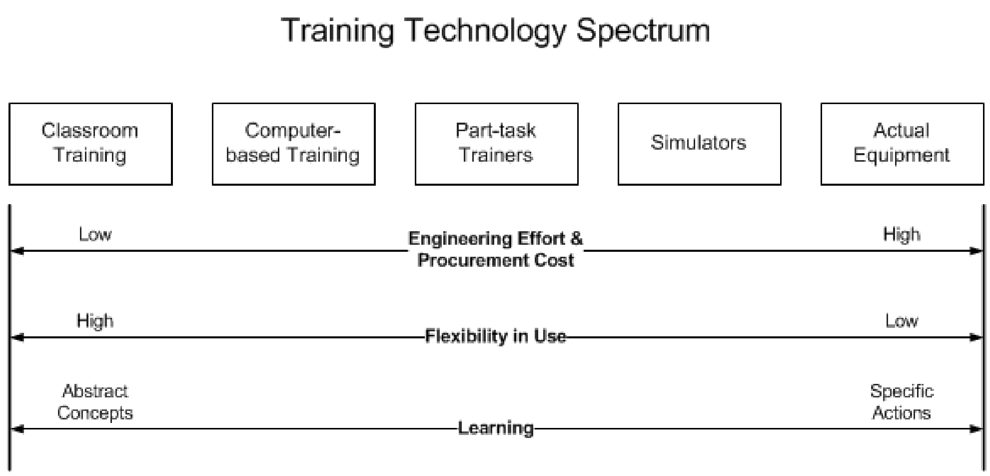


Figure 3. Training technology spectrum (adapted from Ames and Sondhi, 1988).

The choices made by acquisition managers necessitate a compromise between cost, flexibility and learning possible from a given level of training technology. Training on simulators and actual equipment incurs additional costs not only in the expense of the procured hardware, but also in the increased manning levels required to operate and maintain them. Classroom training offers low cost and high flexibility but limits learning capabilities.

While there are several ways of training most skills, the question must be addressed whether a given method is "any good? Does it train as well as, or perhaps, better than some other way we could use. This is the issue of effective training." (Orlansky, et al., 1985). However, the effectiveness of training must be weighed against its cost. Between equally effective training strategies, the clear choice is the cheaper one – similarly, for strategies of similar cost, the more effective is the logical choice. Because candidate strategies are unlikely to be clearly distinguishable, a methodology by which to compare alternatives is needed.

2.2. Learning and Training – a Systems View

Before further developing the relationship between training alternatives and the application of learning models, a distinction must be made regarding the use of the terms *learning* and *training*. Bailey (1989) stated that a "distinction should be made between learning theories and training. Learning theories usually describe the conditions under which a behavior is acquired. Such theories are descriptive." Extending this definition, learning is the process of creating a *general* storehouse of knowledge accessed and applied as necessary to a variety of specific situations. Training theories differ in that they "specify the most effective and efficient ways to obtain knowledge, skills or attitudes at identified levels and under particular conditions. Such theories are prescriptive" (Bailey, 1989). To provide an analog to the extended definition of learning, training creates a *specific* storehouse of knowledge that can be applied to situations that are nearly as specific.

Because this research is framed in the context of training for operators of military systems, using a military operational definition for training is appropriate. The Defense Acquisition Guidebook (DAU, 2006) provides a definition similar to Bailey (1989) with more specificity: training is "the learning process by which personnel individually or collectively acquire or enhance predetermined job-relevant knowledge, skills, and abilities by developing their cognitive, physical, sensory, and team dynamic abilities." Sauer, Hockey and Wastell (2000) expand on this definition by declaring "the goal of training is generally twofold: it should produce rapid skill acquisition among trainees and also lead to high skill retention during periods of non-practice."

Transforming this definition into the context of a systems view of HSI, training can be considered the means by which the human element is adapted or configured to be an effective, contributing component of a system. This systems view can be expanded further by discussing the human abilities (cognitive, physical, and sensory) in the context of software and hardware. Viewed as a machine component, the human is capable of multichannel operation (Aslaksen & Belcher, 1992) because they can:

- accept data input through multiple sensory channels (auditory, visual and tactile senses),
- process that data by means of cognitive skills (thought processing and problem solving), and
- provide output through motor activities.

Essentially, training involves programming these human performance channels in a manner that makes the human component (and overall system) efficient, stable and reliable. While the models described and developed in the remainder of this dissertation are models of training and task practice, the concepts that support learning curves are applicable. Consequently, the terms *learning* and *training* are used interchangeably while discussing characteristics of humans that are learning skills.

2.3. The Learning (Training) Process

The literature offers conceptual models of the learning process. Anderson (1981) proposes a three-stage model of learning that is briefly summarized as:

- *Stage 1: Cognitive learning*. In this stage, the learner is committing a set of problem solving rules to memory as they are introduced to a task. This stage is characterized by slow execution - while a set of rules are in memory, processing is required to interpret and apply them properly.
- *Stage 2: Associative learning*. Links are made between the rules established in Stage 1, resulting in performance improvements in both speed and accuracy.
- *Stage 3: Autonomous phase*. During this stage, the links established during Stage 2 are becoming more automatic. Performance, while still increasing, has less potential for improvement – consequently, proficiency on the task is approaching asymptote.

This process is illustrated in Figure 4. During the learning process, performance is judged primarily by the two most important human performance parameters than can be discussed in quantitative terms: speed and accuracy (Aslaksen et al., 1992). Improvements in these terms are measured as reduced time to complete the task as well as increased accuracy (probability of completing the task successfully).

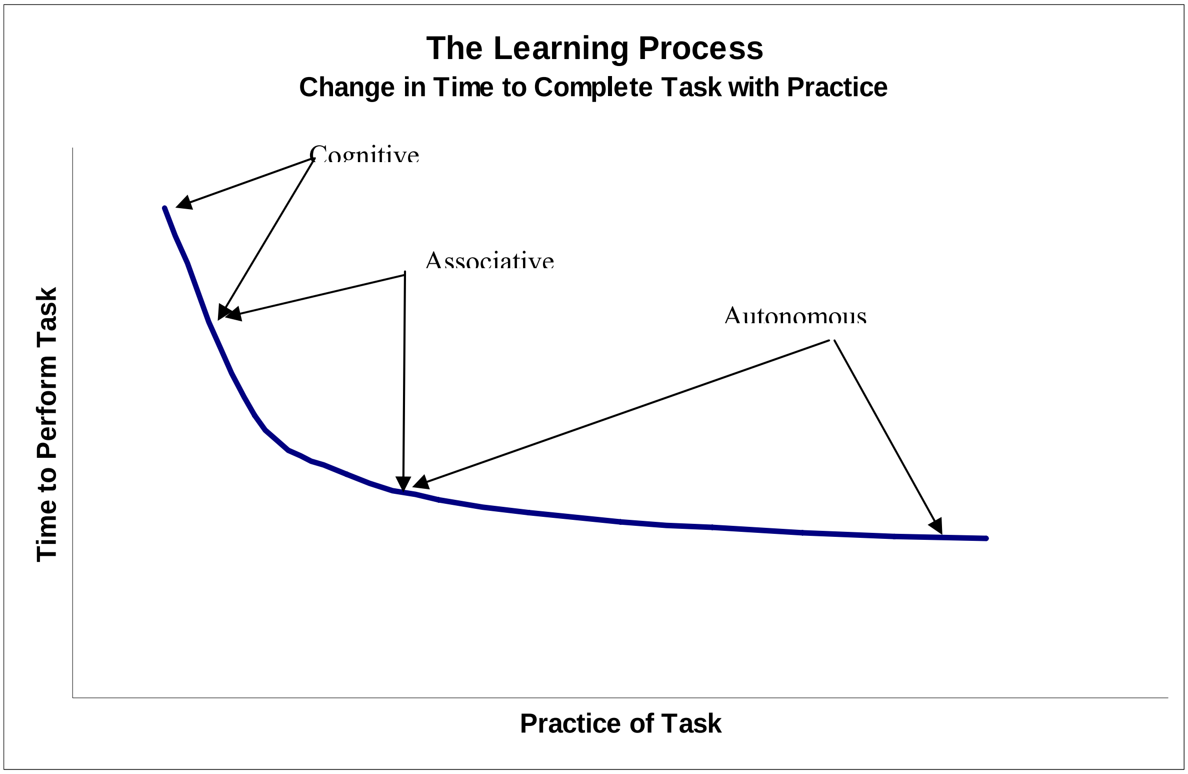


Figure 4. Depiction of learning process

The plot of performance vs. amount of practice in Figure 4 creates a distinct curve – this graphic manifestation of the learning process is routinely referred to as the *learning curve*. The parameterization (and meaning) of the curve's shape is discussed in Section 2.6. However, a discussion of how this curve relates to specific measures of the effectiveness of training – acquisition, retention, and reacquisition of skill – sheds additional light on the uses of a robust training model.

2.3.1. Acquisition of Skill

Assuming that a trainee has little or no prior experience in executing a particular task, the purpose of a formal training program is for the trainee to acquire a specific skill. As illustrated in Figure 5, learning is typically characterized by slow performance during initial attempts with large increments in improvement with each attempt. Performance continues to improve, but with smaller increases with each attempt at the task. At some point in the training process, a performance asymptote is reached suggesting additional practice (trials) has limited effect on improving further performance. Additionally, the magnitude of performance change between the beginning of training and attaining proficiency (asymptote) suggests the overall impact of the training. Characterizing this process and change mathematically helps training managers make informed decisions about the tradeoff between cost and fidelity.

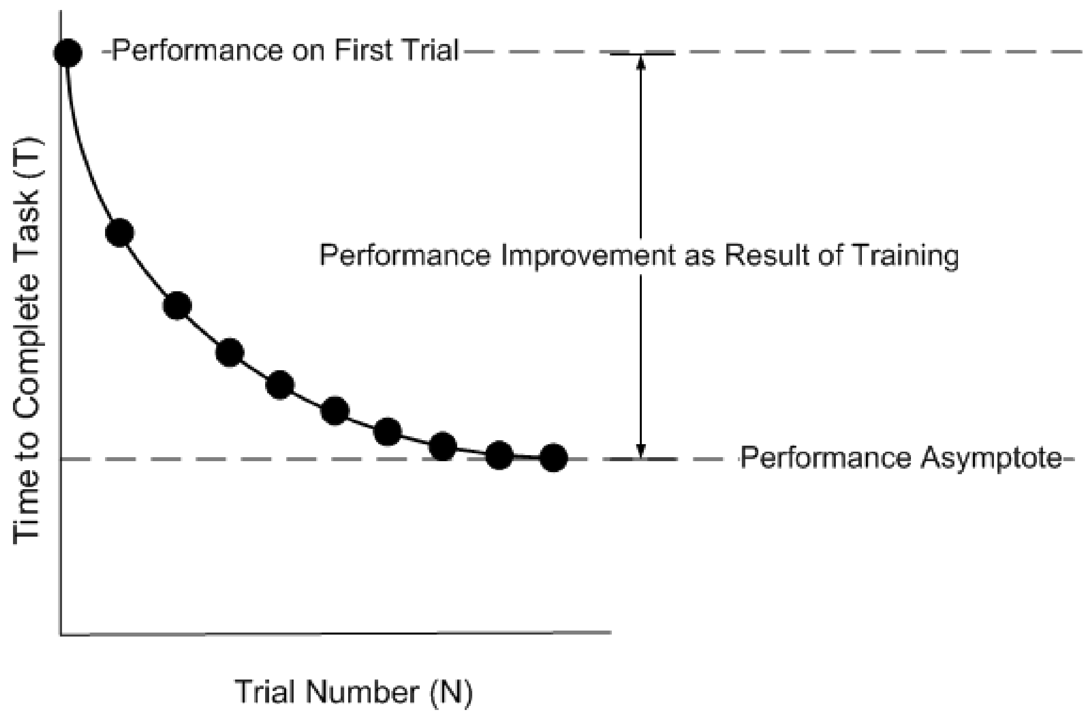


Figure 5. Acquisition of skill curve.

2.3.2. Retention of Skill

Another important measure of the learning process is the amount of the skill retained after a specified period (a retention interval) of nonuse of the skill. Skill retention can be measured by comparing the termination of the acquisition learning curve with the beginning of the reacquisition curve (see Figure 6). Retention of material is often characterized by a "monotonically decreasing function of the retention interval, falling sharply during the time immediately following acquisition, and declining more slowly as additional time passes" (Swezey & Llaneras, 1997). Archer, et al. (2002) concluded that "although a significant amount of what influences decay is determined by acquisition processes, acquisition performance alone is not an accurate predictor of decay." This finding suggests data beyond a skill acquisition curve are required to effectively predict retention.

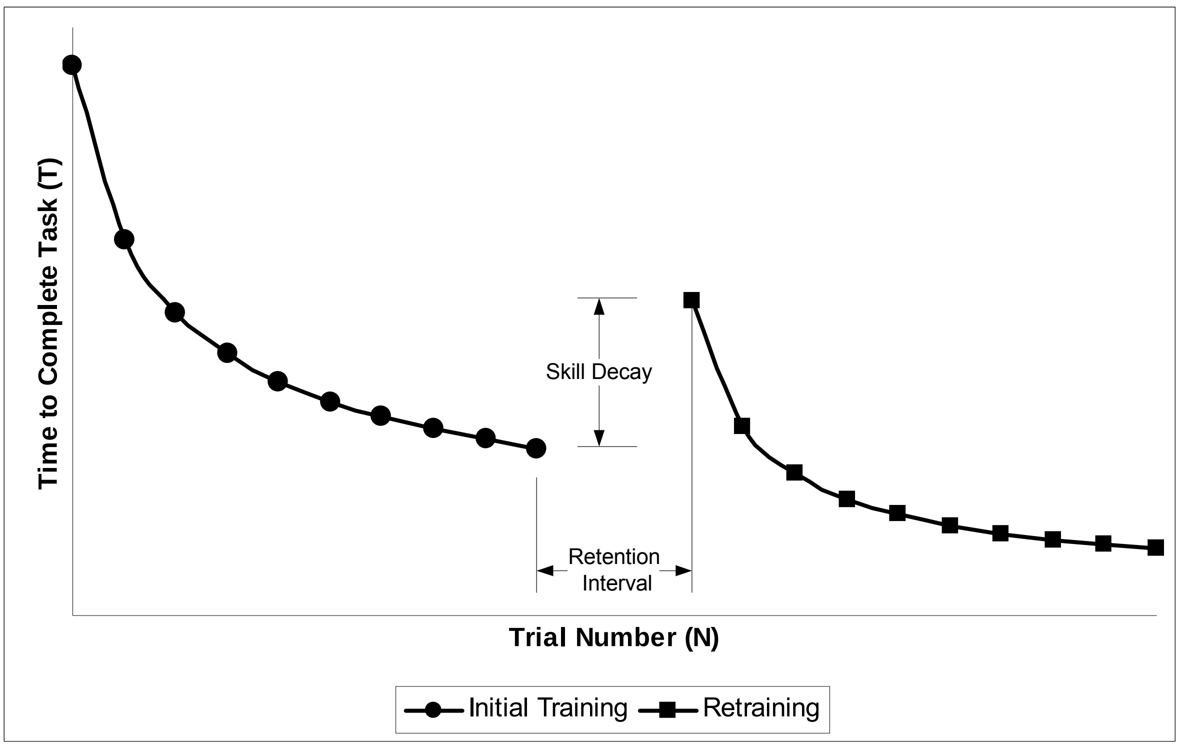


Figure 6. Retention of skill as a function of skill acquisition and reacquisition

2.3.3. Reacquisition of Skill

Closely related to skill retention are the properties of skill reacquisition. If there is indeed a decrease in performance after a period of non-use of a skill (the retention interval in Figure

6), it is logical to assume there must be some period of relearning the task. It is expected that the learning curve for skill reacquisition will have different parameter values (performance improvement, learning rate and asymptote) than the acquisition of skill curve (see Figure 7).

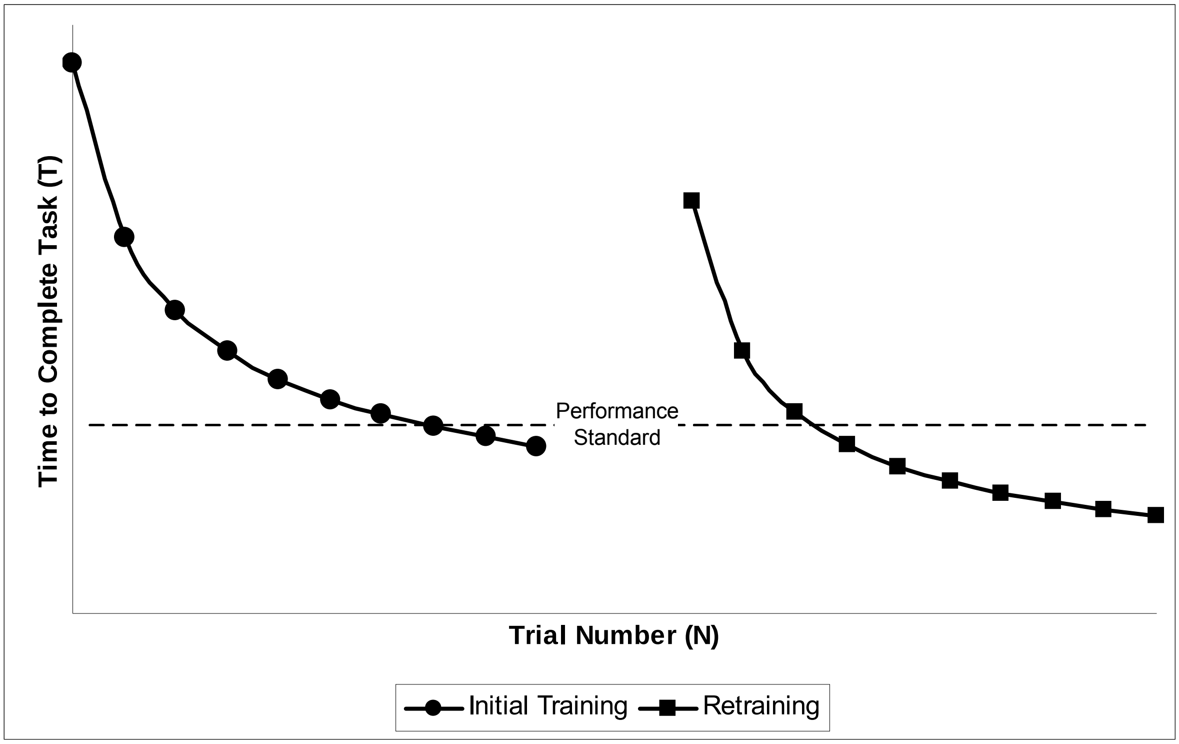


Figure 7. Reacquisition of skill as function of skill acquisition and retention.

2.4. Factors That Impact Training Effectiveness

Past research recommends creating models to measure training effectiveness. However, current models are largely qualitative in nature. The literature describing these models is useful in that they define variables to factor into model development, regardless of whether the model is qualitative or quantitative. Arthur, et al. (1998) performed an extensive literature review and determined the core set of variables include: instructional strategies or training methods, length of the retention interval between initial training and attempts at reaccomplishing a task, and the nature (characteristics) of the task.

2.4.1. Training Strategy

Training strategies are defined as methods through which individuals acquire knowledge, skills and abilities than can be transferred to a real world task (Arthur et al., 1998). Training strategies also include the media through which instruction is given (such as those devices shown in the training technology spectrum in Figure 3). Air Force Manual (AFMAN) 36-2234 (1993) provides several examples of instructional methods, including lecture, demonstration, self-study, computer-based training (CBT) and on-the-job training (OJT). Each of these methods has benefits and limitations and should consider task criticality, difficulty and the instructional fidelity required to teach it.

2.4.2. Retention Interval

Farr (1986) asserts that, in general, the longer the time period of nonuse of a task skill, the greater the decay in the ability to perform that task. Additionally, he suggests that the same factors that affect the ability to initially acquire the skill also affect skill retention.

In a military setting, maximizing skill retention has clear advantages. First, better skill retention reduces refresher training requirements (i.e., the retention interval can be lengthened). Second, improved retention suggests that relearning a task – that is, re-attaining the skill level proficiency gained during initial training – is quicker and requires fewer training resources.

2.4.3. Nature of the Task

According to Farr (1986), the complexity of a task "appears to be the dominant operative characteristic that determines both acquisition and long term retention", citing evidence that complex procedural skills are most apt to suffer from skill decay (he also makes the very important point that tasks performed by military personnel are highly procedural). Because complex procedural skills are ostensibly prone to skill decay because of their difficulty to learn, it is further hypothesized that those types of tasks would also result in a longer period of time to acquire the skill during initial training.

2.5. The Need to Predict Training Effectiveness

According to Allendar (2000), the human component of a system is probably its "noisiest" element. In other words, the human element probably causes the largest variation in system performance. Consequently, predictions of training effectiveness must be sensitive to the variability between humans and the variability humans introduce into system performance. Training is designed to bring out desired performance outcomes; because of the inherent relationship between training and the performance outcome it is necessary to model the cause-effect relationship between training variables and the end result (Salas, Burgess & Cannon-Bowers, 1995).

There are many questions that are of interest to decision makers in the design of training systems. The following sections offer a few examples.

2.5.1. Acquisition of Skill

- Given a choice of training strategies, which provide the greatest change in performance (i.e., the greatest difference between initial attempts at a task and subsequent asymptote (A) at proficiency)?
- Tasks accomplished by military personnel are frequently evaluated according to a standard that is frequently exceeded in both time and accuracy by operators who have attained proficiency in the task. What is the cost tradeoff

between the training strategy (and its associated expense to implement) and the amount of training required to achieve a proficiency standard? Also, is there an interaction between training strategy and the type of task (in terms of what is learned) that must be considered? Figure 8 shows the notional relationship between three training strategies with respect to the speed (measured as the number of attempts) at which a performance standard is attained.

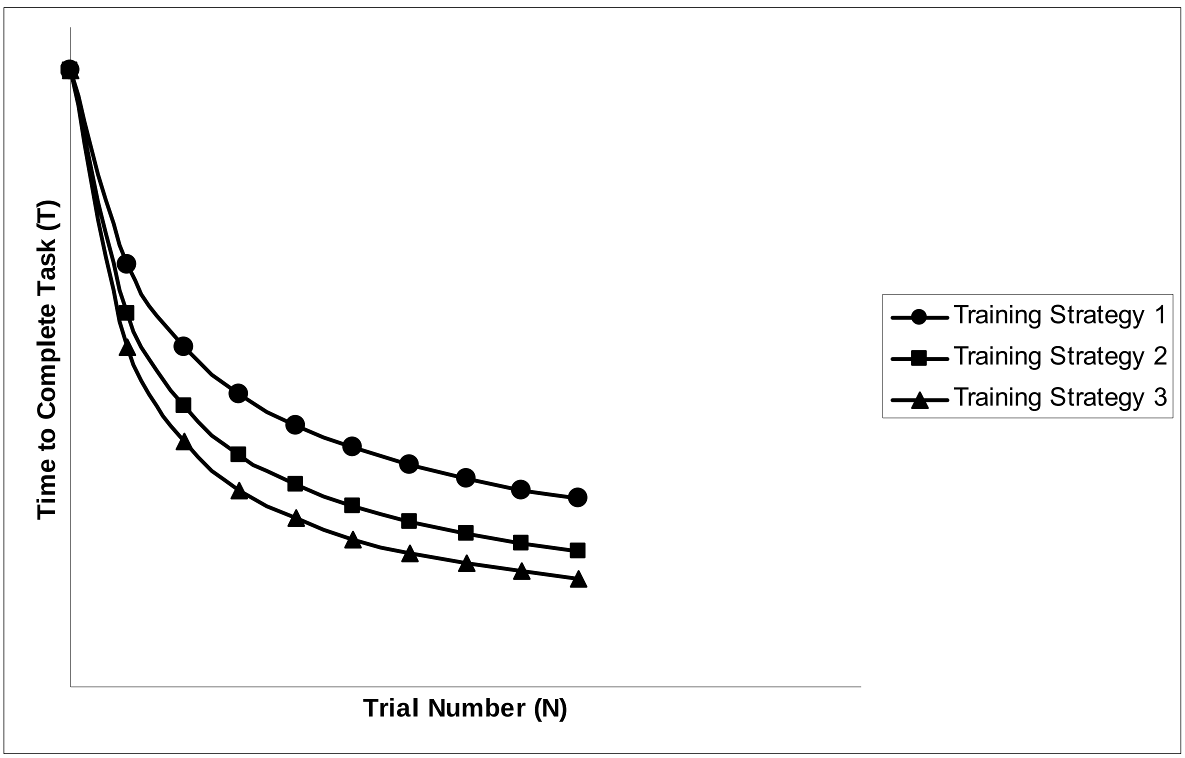


Figure 8. Comparison of training strategies (notional)

2.5.2. Retention of Skill

Modelers trying to estimate human system requirements are interested in addressing questions such as: How quickly (after skill acquisition) should the opportunity to accomplish a task be presented considering the makeup of the task and the method by which it was taught? In other words, how can retention intervals be optimized? Because training opportunities require resources, it would be beneficial to make accurate estimates of the maximum retention interval that still permits performance within defined performance criteria.

2.5.3. Re-acquisition of Skill

Closely related to skill retention are the properties of skill re-acquisition. It is expected that the learning curve for skill re-acquisition will have different parameter values (starting performance, learning rate and asymptote) than the initial skill acquisition curve (see Figure 7). Possible modeler questions are:

- What are the relationships between task type and training strategy in skill re-acquisition?
- For a given training strategy and retention interval, how many times must the skill be practiced before a performance standard is satisfied?
- How long before previous level of proficiency is re-attained?

2.6. Mathematical Characteristics of Human Learning

Attempts to mathematically characterize human learning can be traced back to as early as a doctoral dissertation by Thurstone (1919). Thurstone described the purpose of a learning curve equation as a mathematical method to describe the "law of diminishing returns" – in other words, the notion that unit amounts of performance improvement decrease with practice. Thurstone suggested there would be exceptions to this rule; they may "take the form of a positive acceleration at the initial stage of learning, plateaus during the course of learning and erratic advance of attainment" (attainment suggests a level of proficiency indicated by a stable level of performance).

Thurstone (1919) identified the main parameters of a learning curve as:

- the rate of learning,
- asymptote of learning (the "limit of practice" and point where returns have diminished), and
- an initial learning rate.

Each of these parameters is an important component in the definition of the "family of curves" (Lowry et al., 1992) that have been applied as learning functions used by modelers. Four of the most common family members are described next.

2.6.1. Power Law

Newell and Rosenbloom (1981) are probably the most often-cited on the topic of learning curves and assert that there "exists a ubiquitous quantitative law of practice. It appears to follow a power law . . . We shall refer to this law variously as the log-log linear learning law or the power law of practice." The authors noted that the power function fits a wide variety of skill data, including motor skills, cognitive skills, and perceptual-motor skills. The power law function is distinguished by its property of negative learner acceleration with resulting decreasingly smaller improvements in performance with practice (Swezey & Llaneras, 1997).

The following equation describes the power law in both linear (1) and log-log (2) forms:

$$T = BN^{-R} \tag{1}$$

$$\log(T) = \log(B) - R\log(N) \tag{2}$$

where

T = time to perform a task,

B = performance time on the first trial (N = 1),

N = instance performing the task (trial number), and

R = slope of the line (learning rate).

A graphical depiction of the linear form of the power law is shown in Figure 9.

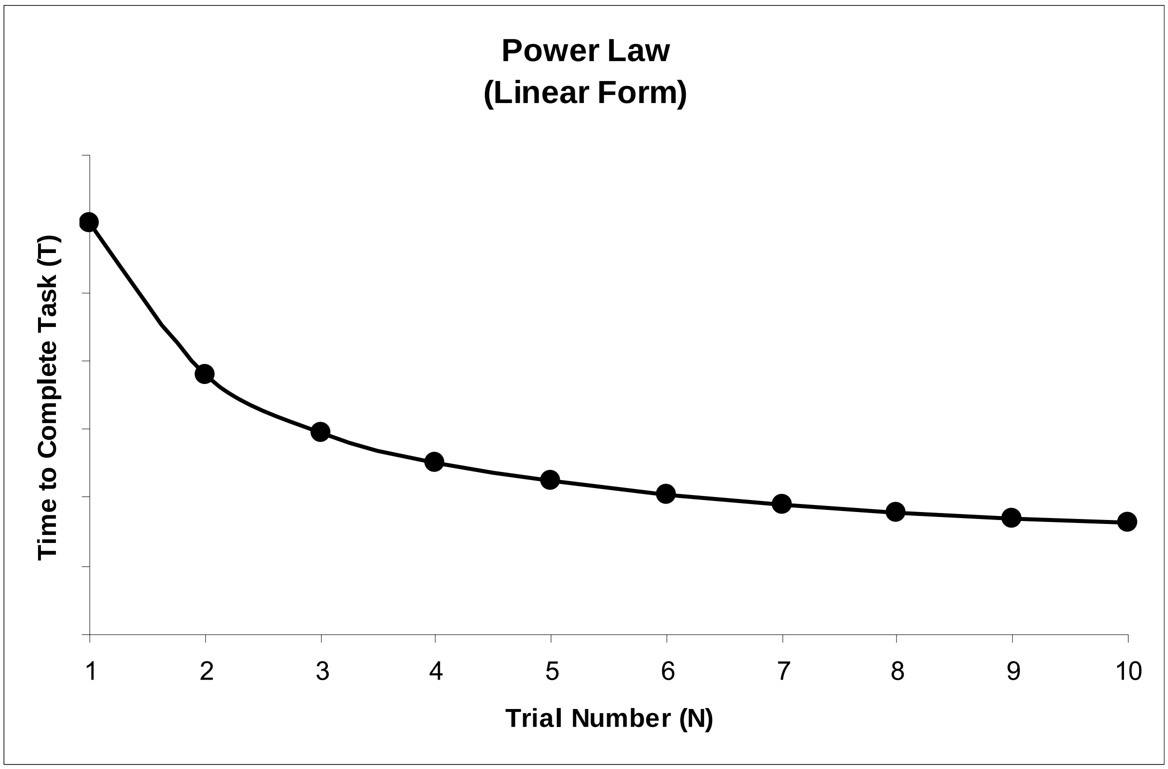


Figure 9. Graphical depiction of the linear form of the power law of learning (time)

Close inspection of Equation 1 reveals a shortcoming that is not immediately evident over a short trial run such as depicted in Figure 9. As the number of attempts to accomplish the task (N) approaches infinity, the time to complete the task (T) will approach zero. Intuition suggests this assumption is false; while performance improvements may yield a very fast performance time, there is some “no better than time” that cannot be improved upon – a performance asymptote. Another assumption that is less evident in Equation 1 concerns the starting value (B) of a sequence of attempts at a task. In this example, it is assumed that all learning begins at N=1. However, the curve as shown in Figure 8 may have been the result of some learning that had taken place before the first trial; some previous experience may have an effect on the curve. Accounting for these assumptions yields a more general form of the power law:

$$T = A + B(N + E)^{-R} \tag{3}$$

where

A = asymptote of learning as N increases indefinitely, and

E = amount of learning (in terms of trials) that occurred prior to the first trial measured.

In more practical terms, asymptote (A) quantifies how good an individual will get, while the term $B(N + E)^{-R}$ provides an indication of the total improvement between the first attempt at a task and the asymptote. All of these parameters are relatively easy to quantify except E (previous experience). It is quite difficult to relate previous knowledge to the performance of a task, particularly so in the units described (number of trials). Consequently, throughout this dissertation, a modified general form of the power law is used (Eq. 4).

$$T = A + BN^{-R} \tag{4}$$

In this form of the equation, the term BN^{-R} indicates the total amount of performance improvement achievable with practice.

2.6.2. Exponential Law

Heathcote and Brown (2000) challenged the notion that the learning curve is best characterized by a power function. They criticized a large body of research that applied the power law of practice without proper tests that may prove another function may be better fit to learning data. The authors suggest the exponential function more accurately reflects the mathematical nature of learning.

The exponential curve differs from the power function in that only one variable is transformed logarithmically (Lowry et al., 1992), and it produces a steeper curve. The form of the exponential function is:

$$T = A + Be^{-RN} \tag{5}$$

where

T = time to perform a task,

A = asymptote of learning as N increases indefinitely,

B = performance time on the first trial (N = 1),

e = natural logarithm,

N = instance performing the task (trial number), and

R = slope of the line (learning rate).

A graphical depiction of the exponential law of learning is shown in Figure 10.

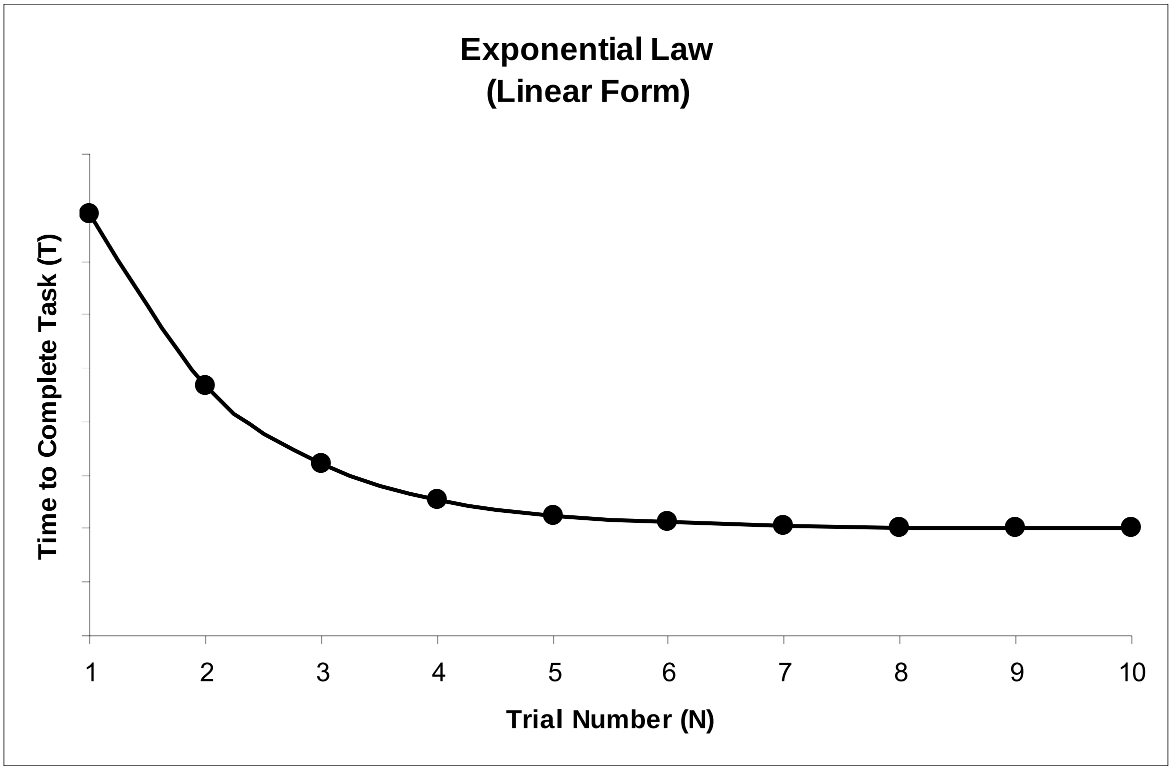


Figure 10. Graphical depiction of exponential law of learning (time)

2.6.3. Hyperbolic

The hyperbolic function was introduced by Thurstone (1919) and is parameterized in much the same way as the power function. There is an important exception – a learning rate of -1 is assumed and is asserted invariant among individuals. This form is characterized by the equation:

$$T = A + B/ (N + E) \tag{6}$$

where

T = time to perform a task,

A = asymptote of learning as N increases,

B = performance time on the first trial (N = 1),

N = instance performing the task (trial number),

E = prior experience.

A graphical depiction of the hyperbolic law of learning is shown in Figure 11.

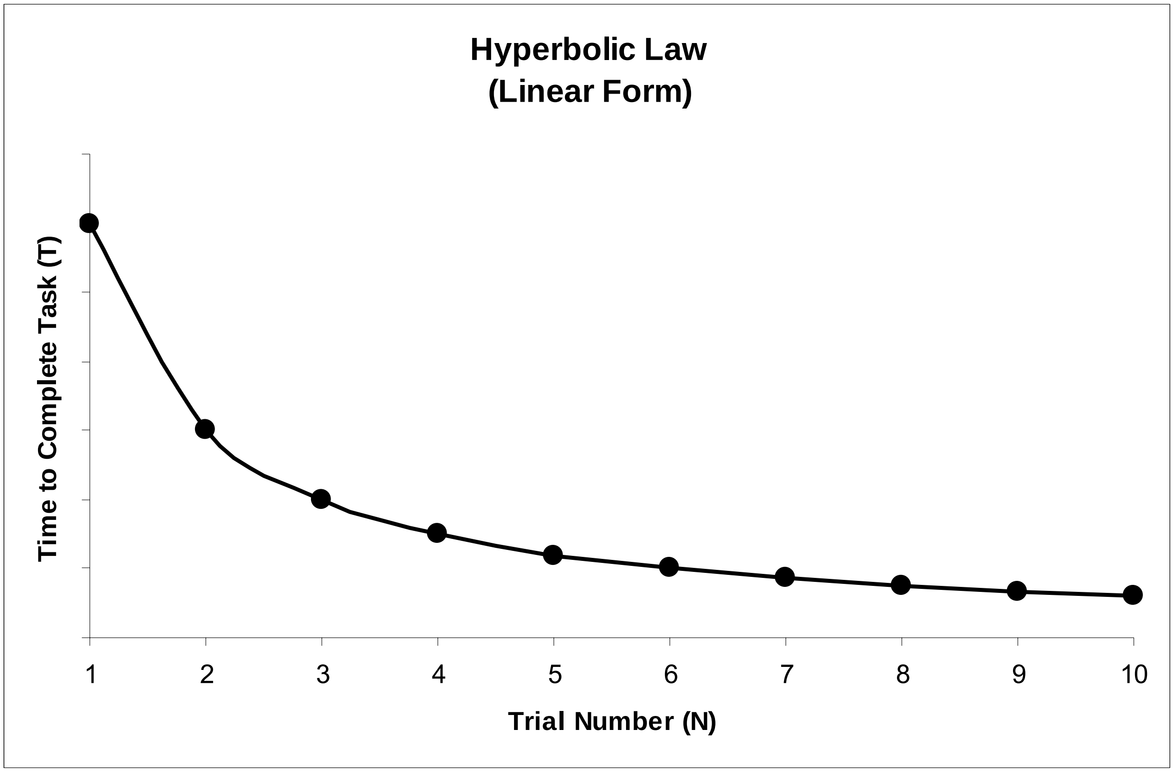


Figure 11. Graphical depiction of hyperbolic law of learning (time).

2.6.4. Logistic

The logistic function is used less than those just described and appears to have been derived largely from subject performance data (Lowry et al., 1992).

The logistic function is characterized by the equation:

$$T = A / (1 + (B - A)\, e^{-kN}) \tag{7}$$

where

T = performance score, ratings, or any other measure that is represented by increasing values,

A = asymptote of learning as N increases,

B = performance time on the first trial (N = 1),

N = instance performing the task (trial number),

e = natural logarithm, and

k = an implicit function of R (learning rate) representing the percentage of learning accomplished on each trial.

Because its construction does not address the types of data collected in this research, Equation 7 will not be considered further.

2.7. Search for a Practical Model of Human Learning

A review of the literature yields few examples of learning models applied to real-world situations. The U.S. Army Research Institute performed a three-year study in an attempt to build a tool that predicted skill acquisition and retention (Rose et al., 1985). This tool, called the unit decision aid (UDA), was a sequence of multiple regression analyses designed to find significant effects between individual soldier characteristics and task performance. Designed to be sensitive to task characteristics and other moderating effects such as soldier ability (characterized by testing scores, etc.) and practice, the tool showed high correlation between predictions and actual learning retention displayed by soldiers in a specific task. The researchers acknowledge the importance of, but did not account for, variables in the modeling of acquisition and retention, particularly training strategy and degree of initial learning.

Lane (1987) performed an analysis to look for relationships between training variables and performance. Findings of this research were (1) while learning curve shapes exhibited some consistency, their parameters varied significantly and could not be relied upon to offer accurate prediction, and (2) no usable patterns in the parameters were apparent.

Lowry et al. (1992) studied the possibility of developing a model that predicted the impact of training variables on human performance, referred to as a Human Reliability Growth Model (HRGM). The HRGM model has three characteristics. First, it provides the capability to predict the effects of training variables on military personnel performance using a mathematical characterization of a learning curve. Second, it is based on empirical data. Finally, it uses taxa-based relationships to determine training impacts as a function of task decomposition (in this context, taxa-based suggests the nature of the task as has been previously described).

Lowry et al. (1992) attempted to develop an HRGM using empirical data collected from an expansive literature review on human performance related to training. Of approximately 3000 research documents reviewed, only 27 documents were found to contain information that would support the construction of an HRGM. The authors "concluded that, although a theoretical basis for development of an HRGM exists, the data could not support its development."

In an attempt to build on the knowledge on HRGM construction, Archer et al. (2002) modeled the effects of training variables across a range of tasks performed by Army soldiers. Their extensive literature review uncovered a wealth of literature on commonly accepted characterizations of training, but found no models to predict the effects of training. In an attempt to bridge the gap between notional and useful models of training, Archer et al. (2002)

created training models for use in simulation software. Unfortunately, the data used to create the models were almost entirely subjective; with no access to empirical data, models were formed using data obtained from performance estimates made by subject matter experts (SMEs).

These findings suggest a learning model, if quantitatively derived, would require experiments that would provide data across a range of the training effects. Ideally, these effects would cover the range of variables of interest to the training designer, such as strategy by which tasks are trained, time intervals between initial training and recall of the task, and the task type. An empirically-derived model built from such experiments requires data processing construction methods that would make it capable of moderating performance (a PMF) across a broad range of instruction methods, retention periods and types of tasks. Additionally, this PMF would be capable of modifying system performance in a way that replicates the variability of that performance attributed to humans. The lack of a model that simulates this variability in simulation tools, IMPRINT Pro in particular, is the subject of the research described next.

3. RESEARCH FRAMEWORK FOR TAXA-BASED MODELS OF TRAINING EFFECTS

3.1. Overview

There is a relatively small body of work in the area of modeling training effects as a function of task composition. While there have been significant efforts to create models of human learning in a conceptual framework, there are few practical models of learning useful as a predictive tool at a higher level. This finding is surprising in light of past assertions made on this topic. For example, Newell and Card (1985) recommended analyzing system design alternatives using predictive, quantitative modeling techniques. Lane (1987) describes the need for better methods to select efficient and effective training methods appropriate to the training situation (the task to be learned). However, despite the recognition that this research area demands further study, this topic is open for exploration and offers vast opportunities for research.

3.2. Research Objective

The objective of this research is to design, implement and evaluate a modeling methodology that will enable a modeler to make training effect predictions. This process will be evaluated by creating a PMF using a set of three tasks that isolate each of one of the three taxon categories (explained in Section 5.1), then applying the PMF to a fourth task for which empirical data are available.

The methodology considers several factors that have not been previously addressed in an integrated way. The following sections provide the general framework for this methodology

– more detailed information on how the methodology was used with real data is contained in Section 5.

3.2.1. Collecting individual performance data and fitting learning curves

Empirical data collected from a population of learners is necessary to create a robust model of learning performance. Data were collected while learners performed tasks that isolated, as much as possible, the perception, cognitive and motor taxon categories. Additional data were collected on a more complex task that combined learning from all three taxon categories. These data were fitted to power curves, providing values for the core parameters of the curve; total performance improvement (B), learning rate (R) and asymptote (A). These values were then analyzed to determine whether training strategy, retention interval or task type have a statistically significant effect on them.

The power function was selected from the family of curves described in Section 2.6. As discussed, the issue of which function best describes learning is controversial and far from being resolved (Cousineau, Helie & Lefebvre, 2003). However, there is a significant body of literature that supports the use of the power curve in modeling research. Swezey and Llaneras (1997) assert that even though the power law has generally been applied to perceptual and motor skills, it appears relatable to cognitive tasks. Newell and Rosenbloom (1981) found particular value of the learning rate parameter of the power function in characterizing learning in a variety of contexts.

This research addresses modeling of the core parameters of the learning function. Because the two most influential functions in learning research, the power and exponential, employ similar parameters, it is asserted that a valid methodology could be applied to both functions as desired.

3.2.2. Test relationships between the core parameters

Intuitively, the three core parameters of the power curve indicate three characteristics of the learning process:

- Performance improvement (B) – the total change in performance as a result of learning.
- Asymptote (A) – the level at which the learner attains consistency in performance.
- Learning rate (R) – the shape of the path between B and A.

From an analytical perspective, the independent variables' (task type, training strategy and retention interval) impact on task performance as manifested by the core parameters is a cascade of increasingly more complex potential effects and interactions (Figure 12). Three analyses are conducted to test for these effects. The first analysis is accomplished on the initial core parameter of performance improvement (B). The next analysis is performed on the learning rate (R) which is tested for the effects of performance improvement (B) as well as the independent variables. Finally, because the asymptote (A) may be influenced by learning rate as well as performance improvement, it is tested for effects of those parameters as well as the independent variables

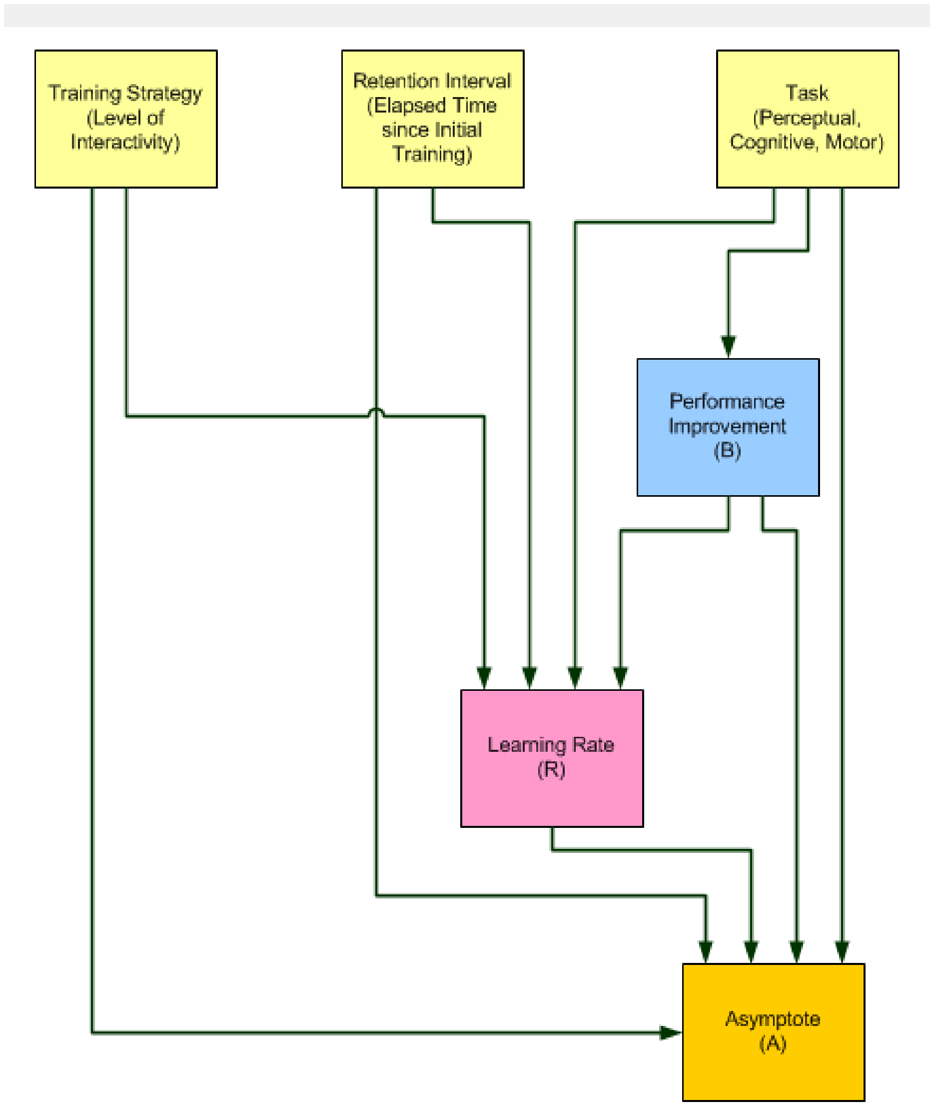


Figure 12. Conceptual framework for modeling process used in this research

3.2.3. Fit core parameters to distribution models

The performance moderators currently installed in IMPRINT are simple algorithms that adjust task time and accuracy identically for all runs in a simulation. Consequently, the PMFs available cannot account for the randomness inherent in human-system performance. By fitting the core parameters to distribution models, the overall training model should predict not only the mean performance for a treatment level but should closely predict the variance as well.

3.2.4. Test models against empirical data

Once constructed, the models are tested against empirical data. The models are tested in two ways. First, the models built from data collected during imagery, TOLD and laser designation tasks are tested against actual data to validate the modeling approach. Next, the models are used to make predictions for a talk-on task that was not included in the model development. A positive fit of the model for those data indicate the models developed are applicable to tasks other than those for which data was collected. These tasks are described in the next section.

4. METHODS

The data collection for this research was accomplished by the researcher for the Science Applications International Corporation (SAIC) in support of the IMPRINT Training Algorithms Enhancement project funded by the Air Force Research Laboratory's Warfighter Reading Research Division (AFRL/RHAT).

4.1. Participants

Sixty volunteers participated in the study. The participants were mostly undergraduate college students. The participants were selected from the undergraduate student population because of similarities to Predator sensor operators (SO) in terms of age, gender, education and operational experience prior to entering training. The graphs in Figure 13 describe the demographics of the participant population by gender, age, years of college completed and whether they had any previous significant aviation or navigation training.

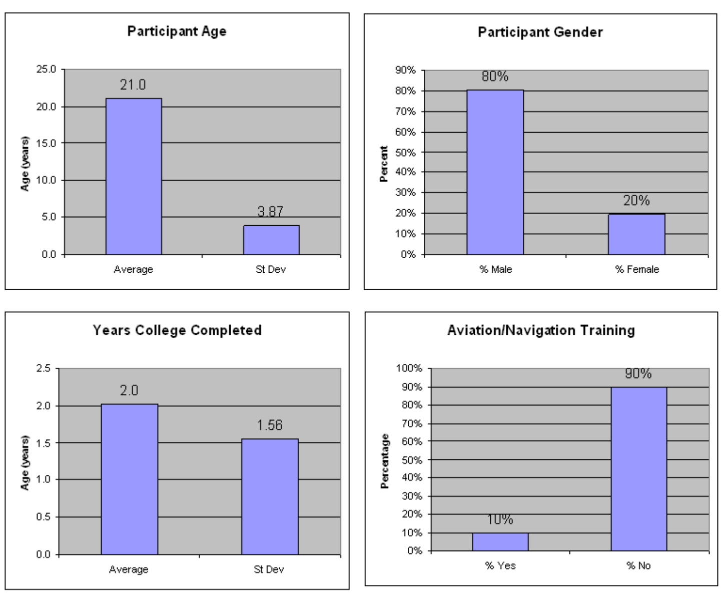


Figure 13. Demographic data for study participants

4.2. Experimental Design and Measures

The study examined participants' ability to perform various sensor operator tasks on an unmanned aircraft system (UAS) sensor simulator as a function of the strategy used to train the participant, the elapsed time since the participant was trained (retention interval), and the nature of the task. The study employed a 3 x 3 x 5 mixed design formed by crossing three levels of training strategy, three retention intervals, and five different sensor operator (SO) tasks. The tasks (described later) were chosen to represent a range of perceptual, cognitive and motor taxons as defined in IMPRINT. Of these five tasks, one was omitted from the analysis because of excessive noise in the data. As analyzed, the study is a 3 x 3 x 4 design.

Ten participants were assigned to each between group treatment level (defined by training strategy and retention interval combination). All 60 participants performed all tasks and all participants completed a training session, an initial data collection session, and a retention data collection session. Between group measures consisted of the training strategy a participant received and whether skill retention was tested at 30 or 60 days. The dependent measure was the time it took to complete the task. Specific details on the training strategies, sensor operator tasks, and how dependent measures were calculated are provided in the following sections.

4.2.1. Training Strategies

All study participants received the same study material delivered by one of three methods according to the training strategy to which they were assigned. Training strategies represented a continuum of interactivity between the participant and the instructor.

4.2.1.1. Low Interactivity

Participants assigned to this strategy completed computer-based training (CBT) modules for each of the experimental tasks. The CBTs were custom-built by SAIC training system specialists according to researcher specifications. The CBTs were completed individually, were self-paced and included graphic and textual animations accompanied by narration. The modules were broken down into lessons that supported specific learning objectives needed to accomplish the tasks. At the end of each lesson, the participant was quizzed on material in that lesson. If a question was answered incorrectly, the correct response and a supporting explanation were provided. At the end of each module, a test covering the entire module was completed. After completing the test, the participant's performance was evaluated, incorrect responses were critiqued and correct answers were provided by the CBT software. The experimenter was available at all times to the participant to answer questions as they arose.

After the orientation and training sessions were completed, participants completed their initial data collection session within two days. As trials were completed, the participants were provided feedback on their performance (whether it was a pass or fail, and whether the fail was attributed to not performing in the allotted time or the trial was incorrectly accomplished). The participants in this training strategy were not coached to improve their performance, but the experimenter would answer any direct questions asked.

4.2.1.2. Moderate Interactivity

The training for this strategy was conducted in a classroom setting with up to six participants in attendance per session. Materials from the CBT were incorporated in a slideshow presentation module for each task. Each module was separated into lessons that supported learning objectives needed to accomplish the tasks. At the end of each lesson, participants completed the CBT quiz for that lesson and received feedback the same as that received by participants in the low interactivity condition. Module tests were also completed and critiqued in the same manner.

In addition to the different setting in which the training information was presented to the participants, the experimenter also provided more elaborate explanations on key concepts. This added emphasis was intended to provide an additional difference between the low and moderate interactivity strategies to represent a differential effect of instructor involvement during training, and reinforce the skills needed to successfully accomplish the tasks.

Participants in this training strategy completed their initial data collection session within two days of the orientation and training session. As trials were completed, the experimenter provided them with feedback on their performance (whether it was a pass or fail, and whether the fail was attributed to not performing in the allotted time or the trial was incorrectly accomplished). If incorrect or inefficient procedures were used during a trial, the experimenter provided guidance to help participants improve performance.

4.2.1.3. High Interactivity

Like the moderate interactivity condition, the training for this strategy was conducted in a classroom setting with up to six participants in attendance. The same materials used in the moderate interactivity strategy were used. The same procedures were used for lesson quizzes and module tests.

To set this training strategy apart from the other two, the experimenter supplemented the concept elaborations by querying the participants to measure their understanding of the material. This additional student-instructor interaction was intended to impart an additional difference between the moderate and high interactivity strategies and further reinforce the skills needed to successfully accomplish the tasks.

Participants in this training strategy completed their initial data collection session within two days of the orientation and training session. As trials were completed, the experimenter provided them with feedback on their performance (whether it was a pass or fail, and whether the fail was attributed to not performing in the allotted time or the trial was incorrectly accomplished). If incorrect or inefficient procedures were used during a trial, the experimenter discussed the trial with the participant and then provided guidance to improve performance.

4.2.2. Retention Intervals

Participants were then assigned to retention interval treatments of either 30 or 60 days. There were three levels for retention interval; 0, 30 or 60 days. All participants belonged to the 0 day treatment level (this was the initial data collection period). Participants were then assigned to retention interval treatments of either 30 or 60 days for the retention data collection period. Because of the logistics involved in participant scheduling, a tolerance of ± 10% of the retention interval was used. After the retention interval, participants performed different instances of all five tasks. The tasks were administered in a random order.

4.3. Facility

The experiment was performed at the SAIC facility in Beavercreek, Ohio. Two areas were prepared for the data collection portion of this study. The participants completed their orientation session and training in a conference room dedicated for that purpose. After completing orientation and training, participants moved to a laboratory to conduct performance trials using the simulator.

Experimental tasks were performed on the Air Force Synthetic Environment for Reconnaissance and Surveillance (AFSERS). The Air Force Modeling and Simulation Resource Repository (AFMSRR) (2008) describes AFSERS as the "primary virtual Intelligence, Surveillance and Reconnaissance (ISR) system used within the Air Force for command and staff level training for the Joint Services." AFSERS is comprised of a generic ground station that can be used to control a variety of UAS and a visualization system used to render imagery from a simulated sensor. For this experiment, AFSERS was used to simulate a UAS sensor operator (SO) station. The simulator was hosted on two PCs. The first PC, referred to as the control station surrogate (CSS), was loaded with several software applications for control of the simulated ground station. The second PC, which represented the SO station, rendered the visual scene using MetaVR's Virtual Reality Scene Generator (VRSG) 3D visualization software. AFSERS version 6.8.1.0 and VRSG 5.2c software were used.

The systems were integrated using a standard network switch and are accessed using standard keyboard and mouse inputs. The sensor was primarily controlled using a standard PC control stick. The control stick allowed the operator to control the pitch and yaw of the sensor and provided the ability to zoom the sensor in and out, control the rate at which the sensor moves, and to designate targets with a laser. See Figures 14 through 16 for illustrations of AFSERS.

Figure 14. The AFSERS system.

Figure 15. AFSERS sensor operator station

Figure 16. Close-up of the sensor operator (S0) station operated by the participants.

A third PC was used as a data collection device. Custom software designed by SAIC software engineers allowed the experimenter to collect time and accuracy data for each trial and save the data directly to a data file. The data file included information such as participant identification number, training strategy used, retention interval, task type, date, and which specific trial instance was attempted. The software also scored participant performance during the laser designation task according to criteria set by the experimenter.

4.4. Procedure

4.4.1. Training Procedures

The academic portion of the experiment described the tasks to the participants and provided the procedural knowledge needed to accomplish them (actual practice was accomplished during data collection on AFSERS). Because of the complexity of the tasks, the academic phase took from three to four hours to complete.

The training was decomposed into five blocks as follows.

4.4.1.1. Background Information

This block introduced the study environment and provided information applicable across the experimental tasks. Features and capabilities of the Predator unmanned air vehicle (UAV) and the role of the sensor operator (SO) were described, followed by a tutorial on navigational headings and coordinate systems (latitude and longitude). Participants were provided information on general terms associated with the aviation and airfield environment, and AFSERS displays and controls were described.

4.4.1.2. TOLD Task

This instructional module provided a brief overview of the importance of TOLD calculations and defined the values the participants would derive. The materials to accomplish the task (mission card, weather information sheet and performance graphs) were described and examples were provided on how to use them. An opportunity to practice problems was included in this module.

4.4.1.3. Talk-on Task

This instruction module started with an overview of what is entailed in talk-on procedures and circumstances under which they are performed. The rest of the module covered techniques for attaining and maintaining directional awareness and the use of unit lengths of measurement as a standard for distance reference.

4.4.1.4. Imagery Task

The imagery task block of instruction provided the participant with techniques to identify key features of imagery in the sensor field-of-view (and vice versa) and to properly maintain directional orientation.

4.4.1.5. Laser Designation Task

This module provided background on how lasers are used to designate targets for attack using precision-guided munitions. Techniques for acquiring and tracking ground targets were also discussed.

4.4.2. Tasks

For each data collection session, the participants performed the tasks in a random order. Each task and a description of how performance was scored on the task are described.

4.4.2.1. Imagery Task

The imagery task required participants to evaluate overhead imagery of a target area and correlate the image with their view through the Predator sensor. Performance during this task was measured as the time required for the participant to declare they see the target and whether they found the correct target.

Participants were provided overhead imagery of selected locations in the continental United States. The imagery typically covered approximately 20 square miles and provided feature resolution similar to a photograph taken from 30,000 feet above ground level (AGL). The image contained several annotations to aid the participant in completing the task. The target was marked on a wide area view and also in a higher resolution inset placed in one of the four corners of the image (information on the size of the inset was provided by a yellow box around the general target area). Arrows pointed to the target in both the wide area view and the inset, accompanied by a brief description of the target. The latitude and longitude coordinates of the target were also provided, but in varying degrees of accuracy to encourage the participants the use that information in finding the target, but without providing all the information needed to find it. An example of the overhead reference imagery for this task is shown in Figure 17.

Figure 17. Sample image from imagery task

As the participants were provided the image, the experimenter reset the AFSERS air vehicle and sensor view to a position near (within approximately 1 nautical mile) the target. When the sensor parked on the selected starting position the experimenter started timing the trial. The participants compared the image with that of the Predator sensor to correlate terrain features and cultural landmarks until they determined the location of the target. When the participant located the target they declared “I have the target” to the experimenter and the trial clock was stopped. The participant then moved the cursor directly over the top of the target (if not already there) to indicate the exact object they have identified as the target. If the participant did not find the target within an elapsed time of 300 seconds, the trial was terminated and was graded as a failure.

4.4.2.2. Laser Designation Task

The laser designation task required participants to move the sensor from a starting position to a defined target and continuously designate the target with a laser. Performance during this task was measured as the time required for the participant to continuously track the target with the laser within 40 feet of the target for 20 total seconds (did not need to be continuous).

The target for this task was the center of the compass rose painted on the dry lake bed north of the runway at Edwards AFB, California (see Figure 18). The experimenter reset each trial by overriding the participant's control of the sensor and moving the sensor field-of-view to the outer circle of the compass rose (approximately ⅓ nautical mile from the target). Upon hearing the experimenter's preparation and execution statement ("Ready, set, go"), the participant actuated the slew rate button on the control stick to release the sensor from its parked position and moved it to the target. Once the sensor was centered on the target, the participant pulled the trigger on the control stick to fire the laser enabling the data collection software to score their performance. The trial was stopped when either the participant maintained the laser designation within the described criteria, or the elapsed trial time had reached 60 seconds, whichever occurred first. If the elapsed time reached 60 seconds the trial was graded a failure. If the participant failed to track the target with the laser for the required duration, the trial outcome was scored as "fail".

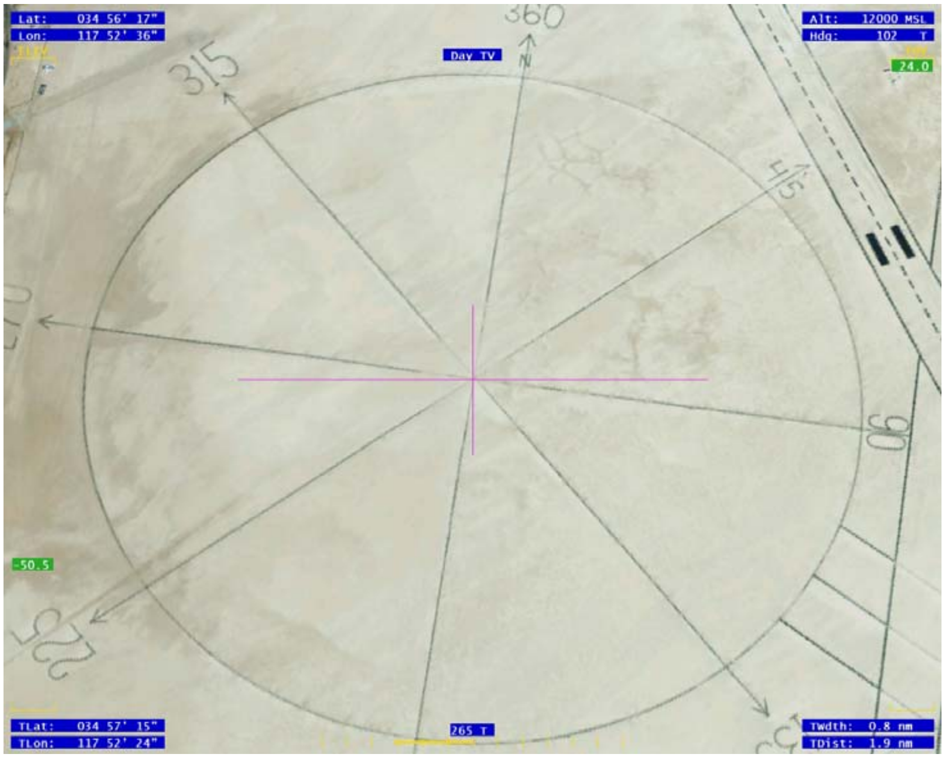


Figure 18. Sensor view for Laser Designation task

4.4.2.3. Talk-on Task

This task required the participant to locate a specific target using information passed verbally from the experimenter. Like the imagery task, performance during the talk-on was measured as the time required for the participant to declare they saw the target and whether they found the correct target. A maximum time of 300 seconds was allowed for each trial; if the participant had not declared they had found a target in that time, or identified the wrong target, the trial was terminated and graded as a failure.

Each trial began with the experimenter confirming the participant was ready to begin. After the participant confirmed they were ready, the experimenter reset the AFSERS air vehicle and sensor view near (within approximately 1 nautical mile) the target. The participant was allowed time to examine the sensor information on the AFSERS head-up display (HUD) to gain awareness of the direction the sensor was pointed. Once they had determined the sensor's direction (or they had elected to skip this step and determined direction in real-time as the trial progressed), the participant stated "ready" and the experimenter began passing talk-on cues and commands.

Three types of cues were provided during each trial – a starting position cue, followed by unit length and directional cues. The starting position cue served to confirm the starting point for the trial. The unit length cue was provided as the distance between two distinct landmarks in the sensor view (or length of a distinctive landmark such as a runway) and was used during the talk-on as a distance reference. Next, the directional cue was given by describing the directional orientation of some landmark(s) in the sensor view and could be used by the participant to either confirm what they had already determined regarding sensor orientation, or it could be used to correct a false notion of direction.

After the cues, movement commands were given. Each movement command contained a direction to move using eight-point cardinal directions (north, south, east, west, northeast, southeast, southwest and northwest), a distance in the form of multiples of unit lengths, and a brief description of the landmark that was the destination of the movement. For example, "Move southwest two unit lengths to the lake." The participant used AFSERS displays and the information received in the cues and movement commands to track the sensor to the landmark indicated. After being commanded to an initial landmark the participant was directed to a second landmark nearby the target. A general target area orientation was provided, followed by a specific target description. When the participant located the target they declared "I have the target" to the experimenter and the trial clock was stopped. The participant then moved the cursor directly over the top of the target (if not already there) to indicate to the experimenter the exact object they identified as the target.

4.4.2.4. Takeoff and Landing Data (TOLD) Task

The “TOLD” task required the participant to perform basic mathematical calculations and derive aircraft performance values from a series of graphs having been provided a table specifying takeoff runway, aircraft weight and weather information.

Six problems were solved in each trial. Two of the problems required the participant to find the headwind and crosswind components present at takeoff. This required them to determine the angular difference between the runway and wind directions and use that information appropriately on a runway winds component chart. A third problem involved determining the maximum allowable crosswind component for takeoff and landing as a function of the aircraft’s weight. The last three problems required the participant to find rotation, liftoff and climb speeds for the Predator as a function of aircraft weight and runway surface condition (wet or dry).

A maximum time of 300 seconds was allowed for each trial; if the participant had not finished the calculations within that time the trial was terminated and graded as a failure. Each of the six problems was checked by the experimenter and evaluated whether their calculation was within two knots of the correct value. If all six problems were correct, the trial was graded as a pass; otherwise the performance on the trial was considered a failure.

5. Results of Data Collection, Processing and Analysis

Before the data was processed, the four tasks used in the analysis were assigned taxon weights according to the procedure described next.

5.1. Taxonomy breakdown of the tasks

As previously discussed, tasks can be characterized by a set of taxons falling into categories for perception, cognition, motor and communication. In current IMPRINT Pro implementation, taxons are weighted in proportion to the amount of time a taxon is called upon during the accomplishment of the task. However, this implementation assumes a proficient operator performing the task. Because this research applies models to the learning process, a different view of taxon assignments is required.

Meador et al., (2008) assert that taxon weights change dynamically during the learning process. Consider again the laser designation task described in the Introduction (Figure 19). This model is an IMPRINT Pro representation of the laser designation task performed by the participants.

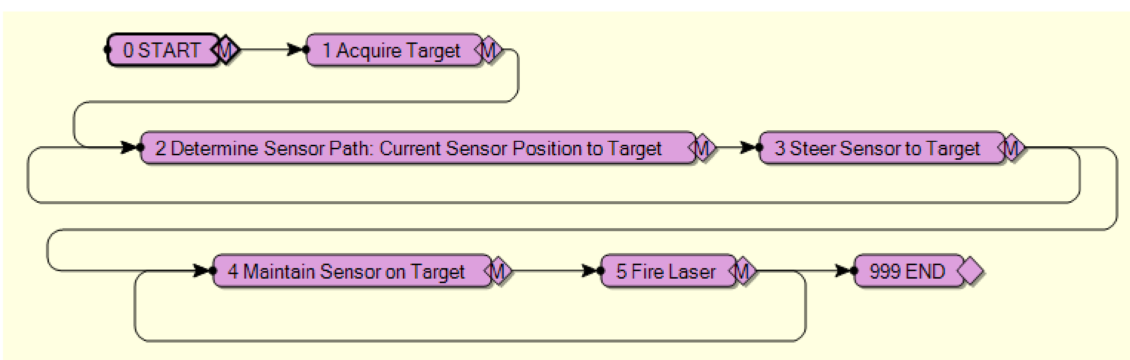


Figure 19. Model of laser designation task performed by the participants.

Table 2 was used to provide an example for taxon weight breakdown for the task when performed by a proficient operator. However, when the task is being learned, the taxon weight breakdown skews the highest weight proportion towards the portion of the task that requires the most practice to attain proficiency. Acquire Target, a perception task, is very easy because the target was plainly in view when the task began; consequently, the amount of time required to accomplish this part of the task remains constant during the learning process. Determining the sensor path from its current position to the target is a cognitive task because it requires processing information gathered during Acquire Target to reach a conclusion about how to proceed with the remainder of the task. As participants became familiar with the task through practice, there may have been a small decrease in the time required to perform this portion of the task.

The following activities in the sequence of events necessary to accomplish this task was where most learning took place. In contrast to the other tasks the participants accomplished that required only relatively gross movements of the AFSERS sensor, the laser designation task required significant control precision. During the first attempt at this task it was commonly observed that the participants were "experimenting" and "discovering" the feel of the AFSERS sensor control, resulting in longer overall elapsed times to complete the task. Because learning is a change in skill through experience, it becomes clear that the portion of the laser designation that really benefits from practice (learning) is the fine motor, continuous portion. In other words, the human performance channel undergoing change while learning this task is the motor channel.

The other tasks accomplished by the participants isolated the taxon categories that mapped to the other human performance channels. The imagery task, though it arguably had a cognitive and motor element, saw improvement primarily as a function of the visual discrimination and recognition taxon (perception). The TOLD task contained elements of numerical analysis (cognitive) and reading and writing (communication). However, it seems illogical to assume that those human performance channels were being trained during practice of this task; after all, the subjects were not being taught how to read, write, or do simple math. Instead, what was really being trained was a set of procedures for accomplishing the task, more of a problem solving task (cognitive).

As a result of this treatment of taxon weightings the imagery, TOLD and laser designation tasks were considered to isolate the three human performance channels of perception, cognitive and motor channels, respectively.

The remaining task, the talk-on, does not map to an isolated taxon category and corresponding human performance channel. The perception abilities of the participants improved as they learned to synthesize information available (both visual and verbal) and process that information effectively. While motor skills were not as precise as that required during the laser designation task, blending the ability to smoothly control the sensor in conjunction with perceptual cues and resulting cognitive activities suggests there was significant learning of the motor aspect of this task.

An assumption was made regarding the communication taxons of oral and reading and writing. These taxons, implemented in these tasks, could easily fall under other taxon categories. For example, reading is arguably a combination of perception-cognitive, while writing could be categorized as cognitive-motor-perception (because a constant loop of seeing-thinking-doing is required). The communication-oral taxon could be broken down similarly; talking suggests a considerable motor component (talking is information output), while listening would align closely with perception (information input). The only taxon in the communication category present in the tasks was the oral component of the talk-on task. Because it was largely the perception (listening) and motor (talking) channels that were active when these activities occurred, they will be considered in those taxon categories for the subsequent data analysis and discussion.

The empirical data were processed and individual learning curves constructed for each block of trials (participant completing a task during a retention period). The relationships between the core parameters were tested and models were constructed.

5.2. Data processing

Before any analysis and model specification, the time data needed to be processed and "performance indices" developed to allow meaningful comparisons across task types. Because the nature of the different tasks led to very different task times, a task time performance index was created to allow direct comparisons of training strategy effects on task time across the different task types.

To create this index a baseline performance time across participants and training strategies was calculated for each of the four task types. In real training terms, the baseline time is considered a performance level at which participants are considered as performing to an acceptable standard. This baseline time was determined by calculating asymptote values for all participants and task types. Asymptote values were calculated by averaging the last three trials of a specific task; this presumes that performance was stabilizing during the last three trials. For each task type, asymptotes were ordered and percentile rankings applied. The top 80% of all asymptotes for the respective task types were all considered as performed to an acceptable standard. The worst of the acceptable performance times were used as the baselines.

For each trial, individual performance time was divided by baseline performance time for that task to yield a performance index for that trial. Table 3 lists the median performance times for the tasks. Table 4 provides an example of the conversion made for one of the participant/training strategy/retention interval/task type combinations.

Table 3. Baseline performance times by task

Task	Taxon Category	Baseline Performance Time (seconds)
Imagery	Perceptual	155
TOLD	Cognitive	107
Laser Designation	Motor	49
Talk-on	Perceptual/Cognitive/Motor	246

Table 4. Sample calculation of performance indices (Participant: 21, Training Strategy: Moderate Interactivity, Retention Interval: 0 days, Task: TOLD)

Trial Number	Actual Performance Time (seconds)	Performance Index
1	173.6	1.62
2	114.7	1.07
3	73.9	0.69
4	66.9	0.63
5	52.3	0.49
6	56.7	0.53
7	54.6	0.51
8	55.4	0.52
9	47.9	0.45
10	42.3	0.40

The data were also smoothed to simplify the curve fitting procedure employed. While a general trend of learning was present in nearly all data collected, varying data noise was present. A smoothing procedure using a simple moving average (averaging an observation with the ones immediately preceding and following it) was applied to the data as illustrated in Figure 20.

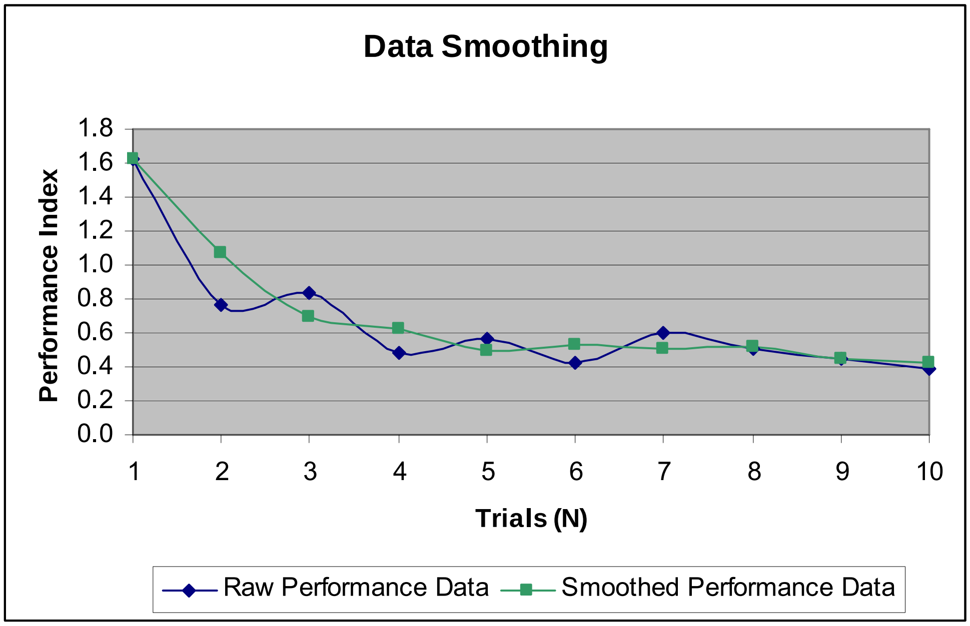


Figure 20. Comparison of raw and smoothed performance data.

5.3. Learning Curve Construction

Learning curves were built for each task block accomplished. These curves were fit to the smoothed data using nonlinear least squares fitting (the method employed by Mathematica®'s nonlinear regression package) and were fit to the power function of learning (Equation 4). Fitting the curve to this function yielded values for performance improvement (B), learning rate (R), and performance asymptote (A). For example, the smoothed data shown in Figure 20 fitted to the model described in Equation 4 yields the expression:

$$T = 0.352 + 1.3*N^{-2.0} \tag{8}$$

Figure 21 illustrates the fit of the Equation 8 function to the smoothed data.

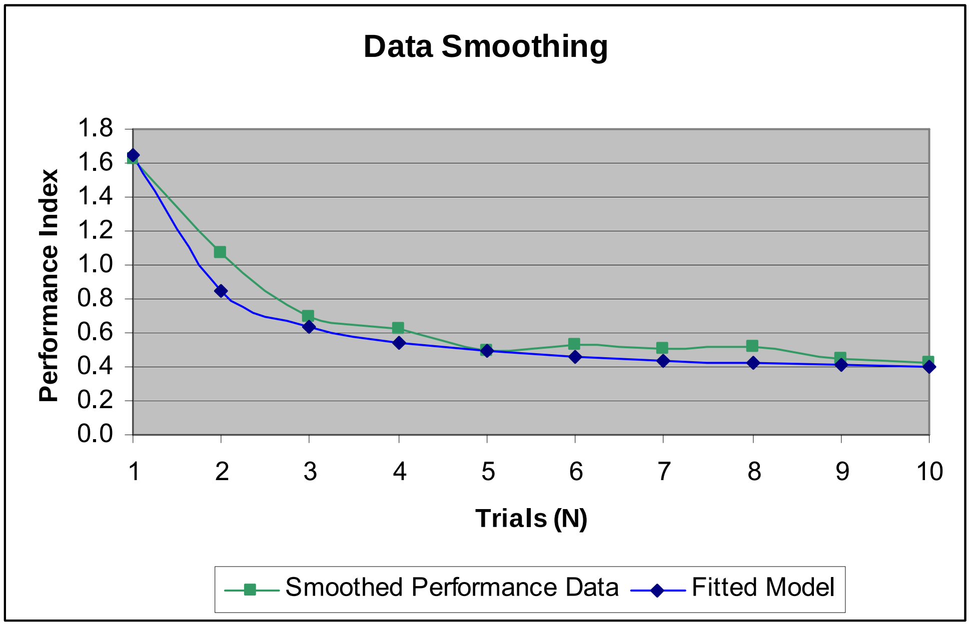


Figure 21. Comparison of smoothed performance data to fitted model.

5.4. Basic analysis on raw performance data

First, an analysis of variance (ANOVA) was performed on the raw data for the imagery, TOLD, and laser designations tasks to test for significant main effects of training strategy, retention interval, task and trial number. Two-, three- and four-way interactions were also tested for significance. Significant levels were set at $p = 0.05$. Results are shown in Table 5.

Table 5. ANOVA for main effects and interactions on performance time

Source	Sum of Sq	DOF	Mean Sq	F	p
TS	3.14	2	1.57	12.98	< .0001
RI	8.09	2	4.05	33.48	< .0001
TASK	55.31	2	27.65	228.84	< .0001
TN	125.60	19	6.61	54.70	< .0001
RI*TS	1.68	4	0.42	3.48	0.0076
TASK*TS	3.80	4	0.95	7.86	< .0001
TN*TS	3.90	38	0.10	0.85	0.7297
RI*TASK	0.45	4	0.11	0.92	0.4499
RI*TN	4.71	22	0.21	1.77	0.0148
TASK*TN	113.20	33	3.43	28.39	< .0001
RI*TASK*TS	1.71	8	0.21	1.77	0.0787
RI*TN*TS	2.89	44	0.07	0.54	0.9939
TASK*TN*TS	3.80	56	0.07	0.56	0.9965
RI*TASK*TN	2.17	38	0.06	0.47	0.9976
RI*TASK*TN*TS	3.60	62	0.06	0.48	0.9998
Error	374.25	3097	0.12		
Total	708.29	3435			

Legend: TS=Training Strategy, RI=Retention Interval, TN=Trial Number

Tukey post hoc tests were conducted to determine differences between treatment levels. The following sections outline the results of that analysis.

5.4.1. Main Effects

All main effects tested had a significant effect on performance. A significant main effect of training strategy ($F(2, 3097)=12.98$, $p<0.00001$) was found between training strategies LOW and MODERATE interactivity. Figure 22 illustrates treatment means and 95% confidence intervals for training strategy.

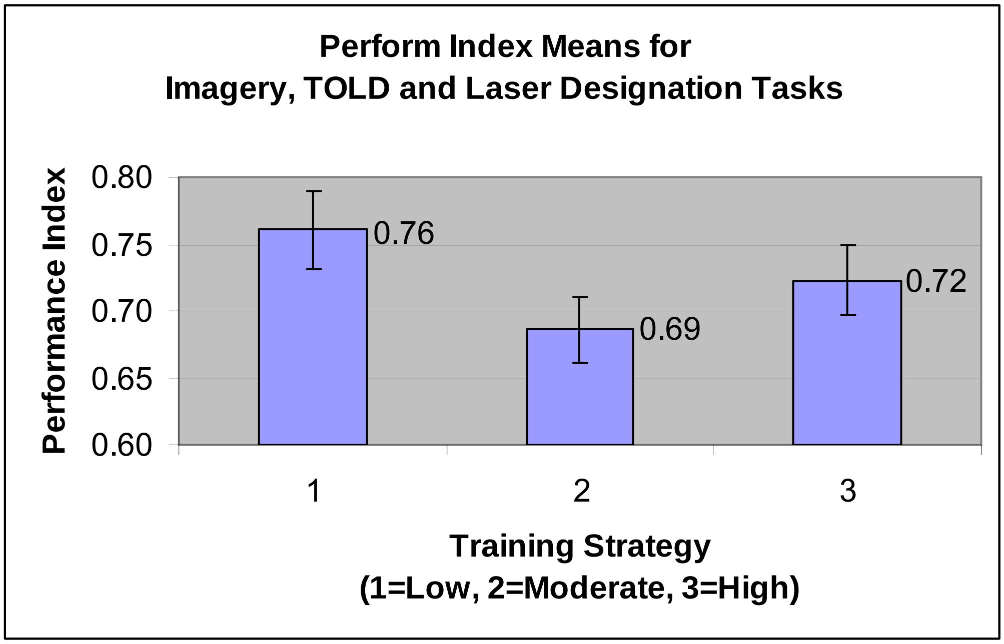


Figure 22. Performance index means and 95% confidence intervals for training strategies.

For retention interval, a significant main effect ($F(2, 3097)=33.48$, $p<0.00001$) was found between the 0 days and 30 and 60 day retention intervals (Figure 23). The 30 and 60 day intervals were not significantly different.

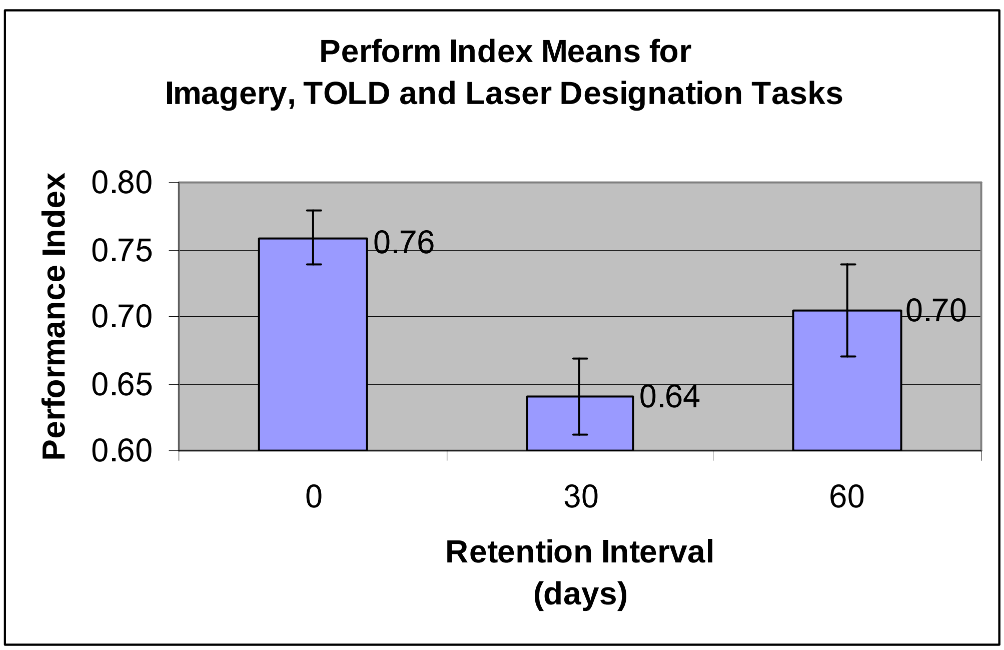


Figure 23. Performance index means and 95% confidence intervals for retention interval.

A significant main effect of task type ($F(3, 3097)=228.84$, $p<0.00001$) was found and indicates that performances on all tasks were significantly different from all of the others (Figure 24).

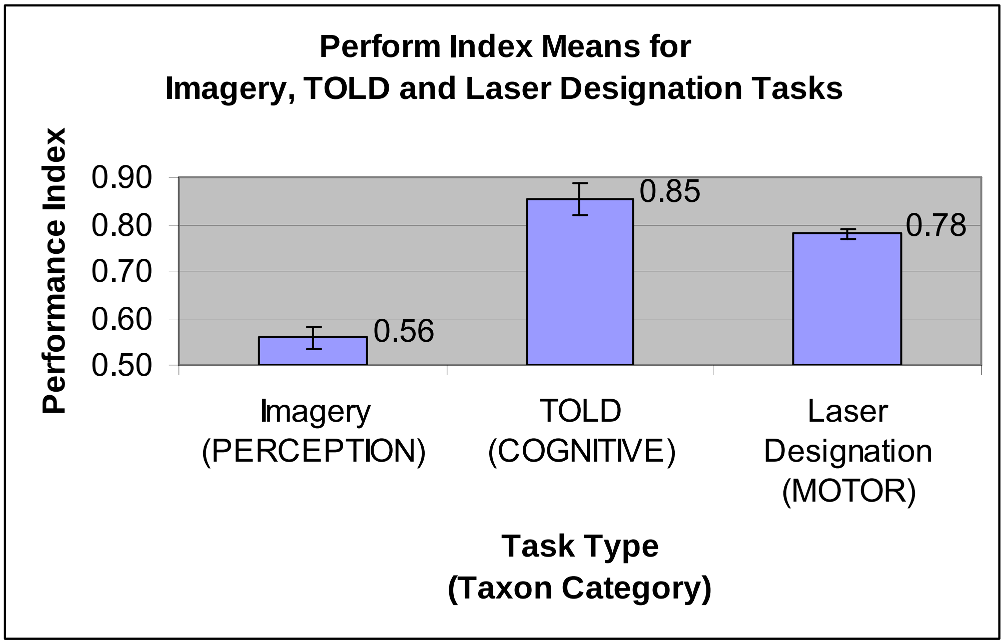


Figure 24. Performance index means and 95% confidence intervals for task type.

Finally, trial number was tested and yielded significant main effects ($F(19, 3097)=54.70$, $p<0.00001$) and generally showed a difference between the trials accomplished early in a task block with those completed as proficiency was attained.

These findings suggest that all independent variables should be represented at some level in a performance moderating function. Because the tasks cannot be included in a grand model, since there is no adequate mechanism to numerically define the imagery, laser designation and TOLD tasks, separate models for each task that account for training strategy and retention interval are warranted. Significant main effects of trial number were analyzed to ensure that learning did indeed take place during each of the tasks.

5.4.2. Interactions

There were several significant interactions on task performance time. Training strategy and retention interval were significant (F(4, 3097)=3.48, p=0.0076), primarily between the low and moderate interactivity training strategies and the initial training period (RI=0) and the 30 and 60 day retention intervals. See Figure 25 for treatment level means and 95% confidence intervals for this interaction.

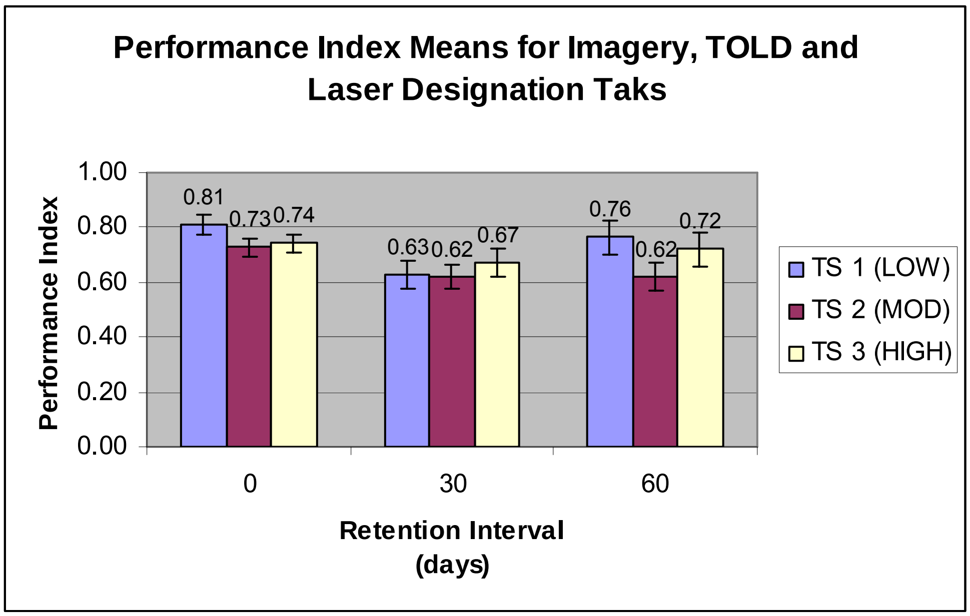


Figure 25. Treatment level means for the training strategy and retention interval interaction on performance time.

There was a significant interaction of training strategy and task (F(4, 3097)=7.86, p<0.00001). Within training strategies 1 and 3 (low and high interactivity), there were several differences between the task types. Figure 26 illustrate interaction means and 95% confidence intervals.

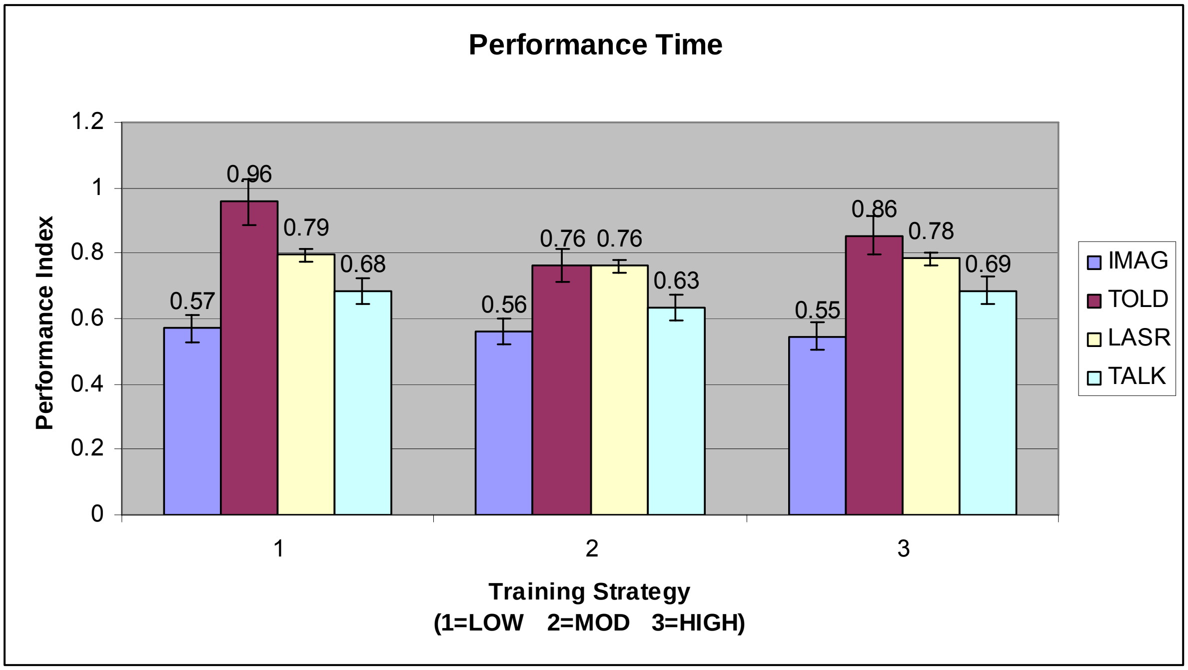


Figure 26. Treatment level means for training strategy and task type interaction on performance index.

5.5. Analysis of effects on and between core parameters

While the previous analysis indicates there was a significant main effect of task type on performance, it is not feasible to include task type as a term in a training effect model since the tasks do not lend themselves to a numerical coding required to include them as a model term. Consequently, separate models for performance are built for each of the task types.

Analyses were conducted on the significant main effects and interactions on the core parameters of performance improvement (B), learning rate (R), and asymptote (A). Tables 6 through 8 list the data associated with ANOVAs for those dependent variables.

Table 6. ANOVA for significant main effects and interactions on performance improvement (B)

Source	Sum of Sq	DOF	Mean Sq	F	p
TS	0.34	2	0.17	0.768	0.4648
RI	4.44	2	2.22	9.951	< .0001
TASK	116.43	2	58.22	261.123	< .0001
RI TS	0.31	4	0.08	0.346	0.8466
TASK TS	1.82	4	0.45	2.041	0.0885
RI TASK	1.70	4	0.42	1.905	0.1093
RI TASK TS	0.96	8	0.12	0.538	0.8279
Error	71.57	321	0.22		
Total	197.56	347			

Legend: TS=Training Strategy, RI=Retention Interval

The only significant main effects or interaction on performance improvement (B) was for task type (F(2, 321)=261.12, p<0.0001) and retention interval (F(2, 321)=9.95, p<0.0001). All of the tasks were significantly different from the others (Figure 27). The 0 retention interval (initial training) differed from both 30 and 60 day retention intervals (Figure 28).

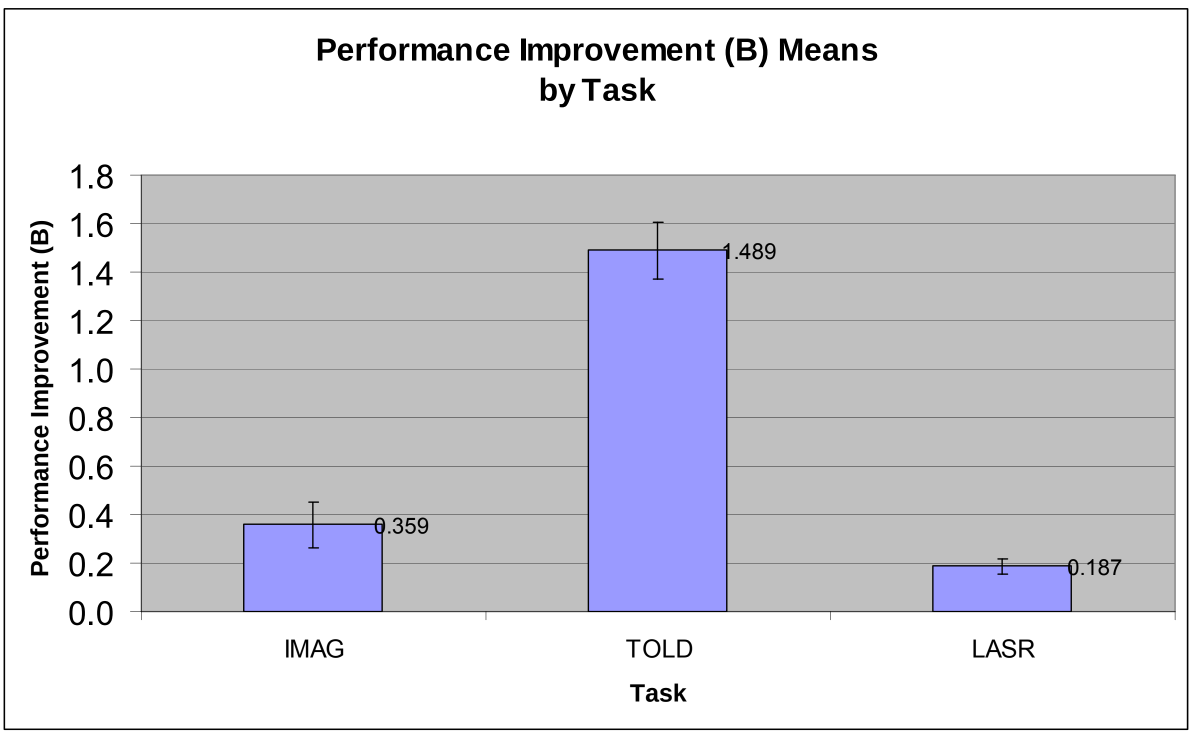


Figure 27. Performance index means and 95% confidence interval by task.

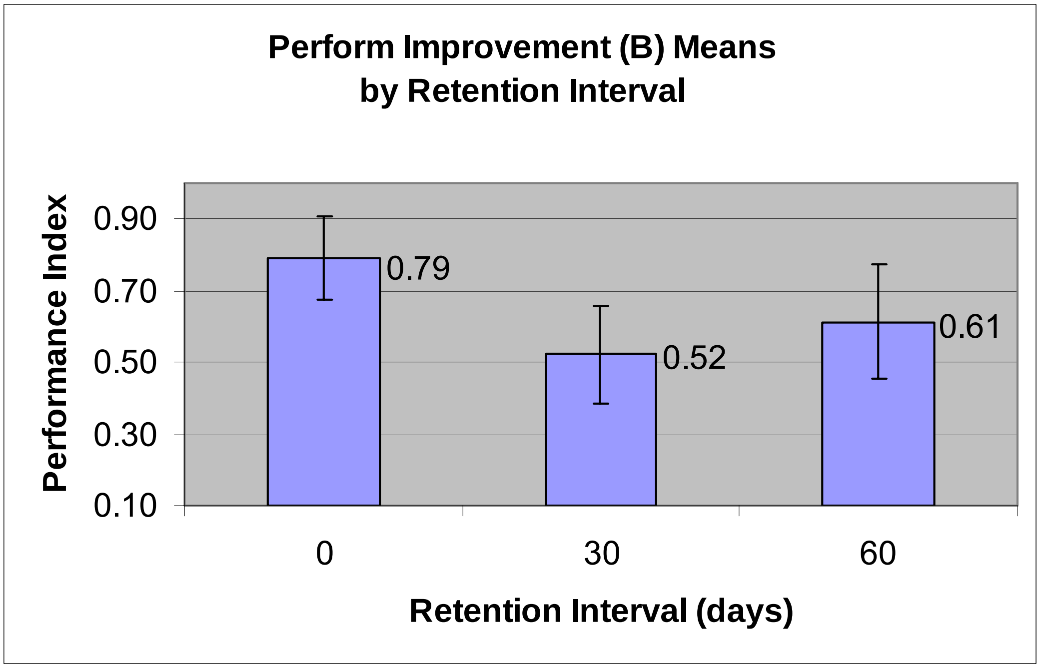


Figure 28. Performance index means and 95% confidence interval by retention interval

The values for performance improvement (B) covered a significant range of continuous values. To complete an effect analysis of the independent variables and B on learning rate (R), reducing the number of degrees of freedom for B was necessary (with all levels of B represented, there were no error degrees of freedom left to accomplish an ANOVA). To accomplish this, values for B were rounded to the nearest .5, effectively creating several bins for B. See Table 7 for the ANOVA for R.

Table 7. ANOVA for significant main effects and interactions on learning rate (R)

Source	Sum of Sq	DOF	Mean Sq	F	p
TS	212.31	2	106.15	2.364	0.0960
RI	922.98	2	461.49	10.279	< .0001
TASK	10923.60	2	5461.82	121.657	< .0001
B	31303.20	5	6260.64	139.450	< .0001
RI TS	133.82	4	33.46	0.745	0.5620
TASK TS	11.72	4	2.93	0.065	0.9921
B TS	52.36	10	5.24	0.117	0.9996
RI TASK	6.75	4	1.69	0.038	0.9973
B RI	55.98	10	5.60	0.125	0.9995
B TASK	2.97	3	0.99	0.022	0.9955
RI TASK TS	352.24	8	44.03	0.981	0.4514
B RI TS	1637.54	17	96.33	2.146	0.0602

B TASK TS	0.19	6	0.03	0.001	1.0000
B RI TASK	0.21	6	0.03	0.001	1.0000
B RI TASK TS	0.04	2	0.02	0.000	0.9996
Error	11762.60	262			
Total	57378.50	347			

Legend: TS=Training Strategy, RI=Retention Interval, B=Performance Improvement

There were several main effects and interactions affecting learning rate (R). The learning rate as a function of task (F(2, 262)=121.66, p<0.0001) was significantly higher for the TOLD task than the imagery and laser designation tasks. Retention interval was significant (F(2, 262)=10.28, p<0.0001) between the retention interval at 0 days and 30 days, indicating higher learning rates during initial training than during retention testing after 30 days. A significant effect of performance improvement (B) was also found (F(5, 262)=139.45, p<0.0001) that indicated a strong correlation between large performance improvements and learning rates.

Similar to the analysis for learning rate (R), the ANOVA for asymptote (A) required binning the continuous values for performance improvement (B) and (R). A resulting large number of degrees of freedom for R (values were rounded to the nearest whole number) precluded the analysis of interactions (Table 8).

Table 8. ANOVA for significant main effects and interactions on asymptote (A)

Source	Sum of Sq	DOF	Mean Sq	F	p
TS	0.18	2	0.09	2.10	.1244
RI	0.44	2	0.22	5.21	.0059
TASK	5.44	2	2.72	64.35	< .0001
B	0.13	4	0.03	0.77	0.5451
R	0.64	9	0.07	1.68	0.0926
Error	12.84	304			
Total	19.66	323			

There were significant main effects for retention interval (F(2, 304)=5.21, p=.0059) and task type (F(2, 304)=64.35, p < .0001) on asymptote (A). For retention interval, there was a significant difference between the 0 and 30 day intervals. Tukey post hoc tests showed that all task types were different from one another.

6. Model Construction, Validation and Analysis

6.1. Model Construction

6.1.1. Model of Performance Improvement (B)

The first component of the overall learning effects model is performance improvement (B). While there is a significant main effect of task type on performance improvement, there is no feasible regression model with which to characterize it. In other words, there is no feasible coding mechanism for task type. Consequently, the best method for modeling B in a simulation as a function of task type is to define a distribution for B that represents the data collected. For the three tasks used to build the model, data for B for each task was fitted to a continuous distribution function for each level of retention interval (the other variable showing an effect on B). Distribution models for each combination of the tasks/retention intervals are listed in Table 9.

Table 9. Distributions for beginning performance (B)

Task	Retention Interval	Distribution Type	Distribution Model
Imagery	0	Lognormal	-1 + LOGN(1.45, 0.643)
Imagery	30	Lognormal	-0.81 + LOGN(1.04, 0.47)
Imagery	60	Lognormal	-0.71 + LOGN(1.04, 0.561)
TOLD	0	Beta	0.06 + 2.77 * BETA(2.36, 1.7)
TOLD	30	Beta	0.21 + 2.28 * BETA(1.12, 1.51)
TOLD	60	Beta	0.19 + 2.52 * BETA(1.69, 1.84)
Laser	0	Normal	NORM(0.242, 0.175)
Laser	30	Normal	NORM(0.159, 0.14)
Laser	60	Normal	NORM(0.107, 0.149)

6.1.2. Model of Learning Rate (R)

According to the ANOVA discussed in Section 5.4, there are significant main effects of retention interval, task, and performance improvement on R. As with the model for B separate distributions for each of the tasks are needed. Consequently, for each task, the

ANOVA suggests 30 distribution models (10 combinations of task type and B x 3 retention intervals). Because such a large number of distributions may result in poor fits, it is necessary to reduce the number of distributions that define R to the minimum amount feasible. Further inspection of the cell means for R reveal that, while statistically significant, the different values for R for the three retention intervals make no practical difference on learning. The values for retention intervals 0, 30 and 60 days reveal learning rates of 1.5, 1.2 and 1.3 respectively. These values are plotted for a constant performance improvement (2.0) and asymptote (0.5) and show the outcomes at these levels of R are virtually the same (Figure 29). Consequently, only values for B are modeled for each of the task types.

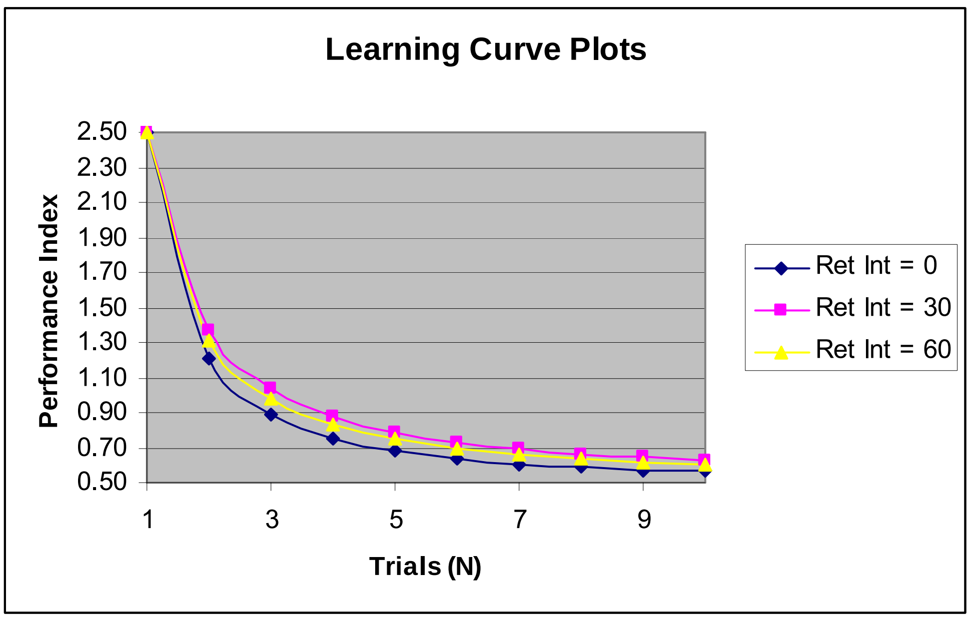


Figure 29. Comparison of statistically significant values of R as function of retention interval.

There are a total of 10 values for combinations of task type and performance improvement (recalling the binning procedure described in Section 5.5). Distribution models for each of the tasks and learning rate combinations are listed in Table 10.

Table 10. Distribution models for interaction between task type and performance improvement (B) on learning rate (R)

Task	Performance Improvement (B)	Distribution Type	Distribution Model
Imagery	0.0	Beta	-1.65 + 3 * BETA(6.09, 2.07)
Imagery	0.5	Beta	1.15 + 0.49 * BETA(0.828, 0.819)
Imagery	1.0	Beta	1.57 + 0.36 * BETA(1.08, 0.782)
Imagery	1.5	Beta	2 + 0.771 * BETA(0.509, 0.753)
TOLD	0.5	Beta	1.27 + 0.36 * BETA(0.961, 1.02)
TOLD	1.0	Beta	1.56 + 0.441 * BETA(1.88, 1.57)
TOLD	1.5	Beta	2 + 0.991 * BETA(0.36, 0.902)
TOLD	2.0	Beta	3 + 11 * BETA(0.585, 1.95)
Laser designation	0.0	Beta	1.21 * BETA(2.47, 1.42)
Laser designation	0.5	Beta	1.06 + 0.48 * BETA(1.25, 1.16)

All of the values for R were fitted to Beta distributions that shared four common parameters; two constants that affect the position of the distribution (K_0 and K_1) and two shape parameters (α_1 and α_1). The form of the distributions are shown in Equation 9.

$$K_0 + K_1{*}BETA(\alpha_{1,}\, \alpha_2) \qquad (9)$$

For each of the Beta distributions the parameters could be fit to models as a function of the value for B. This permits a value for R to be randomly drawn from a distribution for any value of B (not just the binned values as shown in Table 10). A representative plot of the model parameters for the imagery task is shown in Figure 30. All plots and models for Beta distribution parameters for learning rate (R) are provided in Appendix A.

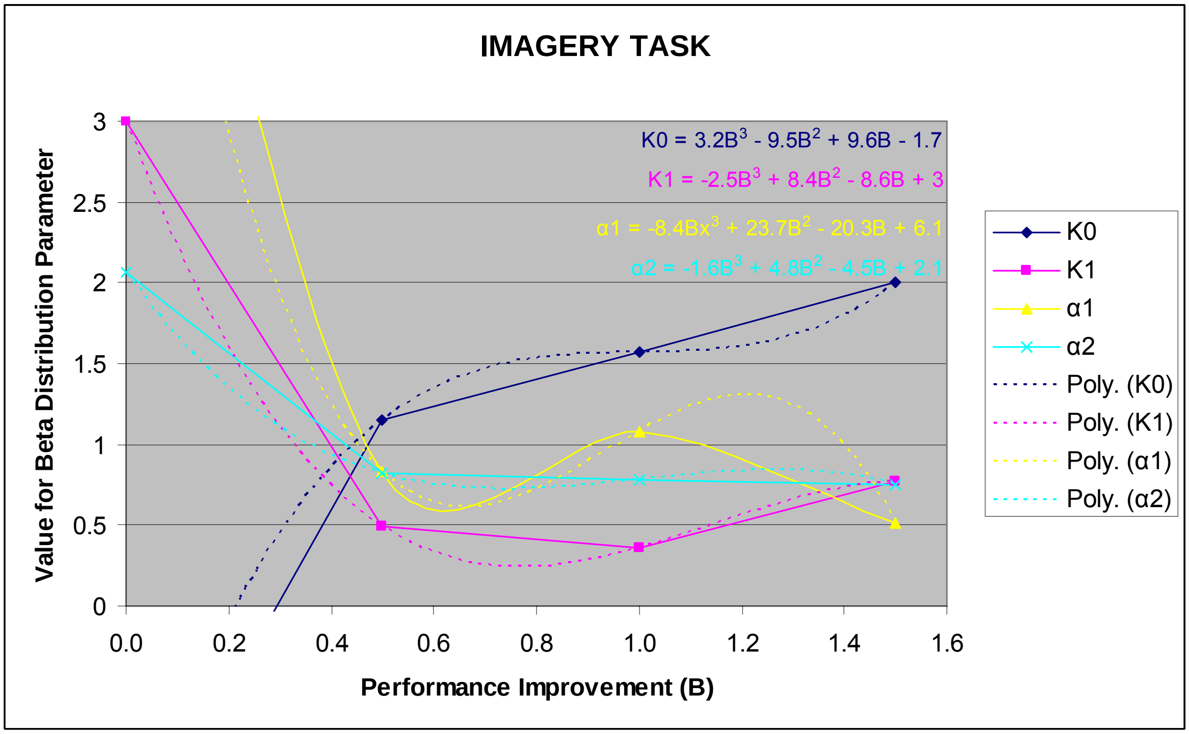


Figure 30. Plot of Beta distribution model parameters for learning rate (R) for the imagery task.

6.1.3. Model of Asymptote (A)

The model for asymptote (A) followed a similar process as that for learning rate (R). The only significant main effects on A were task type, retention interval and learning rate. However, no discernable pattern of A could be inferred from R. The ANOVA results for the significant main effect of R appears to be a result of one or two outlier R values. Consequently, distribution models for A were built for the nine combinations of task and retention interval. See Table 11 for these distributions.

Table 11. Distribution models for interaction between task type and retention interval on asymptote (A)

Task	Retention Interval	Distribution Type	Distribution Model
Imagery	0	Lognormal	0.04 + LOGN(0.439, 0.253)
Imagery	30	Lognormal	0.07 + LOGN(0.288, 0.150)
Imagery	60	Lognormal	0.10 + LOGN(0.304, 0.221)
TOLD	0	Lognormal	0.12 + LOGN(0.475, 0.263)
TOLD	30	Lognormal	0.19 + LOGN(0.309, 0.176)
TOLD	60	Lognormal	0.19 + LOGN(0.446, 0.274)
Laser designation	0	Lognormal	0.52 + LOGN(0.224, 0.159)
Laser designation	30	Lognormal	0.48 + LOGN(0.216, 0.115)
Laser designation	60	Lognormal	0.52 + LOGN(0.244, 0.156)

6.2. Model Validation 1

The first model validation was accomplished by comparing the model to the data that were used to construct it.

6.2.1. Model Execution 1

Executing the model is a multi-step process. First, the task for which data are to be fit is selected. The distribution for that task is then used to provide several values for performance improvement (B). For this example, twenty values were drawn from the distribution, replicating the number of participants' data that was used to create the model.

Next, learning rates for each of the simulated participants were calculated. This step involves using the B values generated and calculating the four Beta distribution parameters as a function of B and the task being simulated. Those parameters are then used to create a distribution for R from which a single value is extracted and paired with the B used to create it.

Next, a value for asymptote (A) is extracted in a manner similar to R; an exception is that A is drawn from a distribution as a function of task type and retention interval. The distribution that was formed as a result of that interaction is used to generate an asymptote for each of the B and R pairings. See Figure 31 for a flowchart of this process.

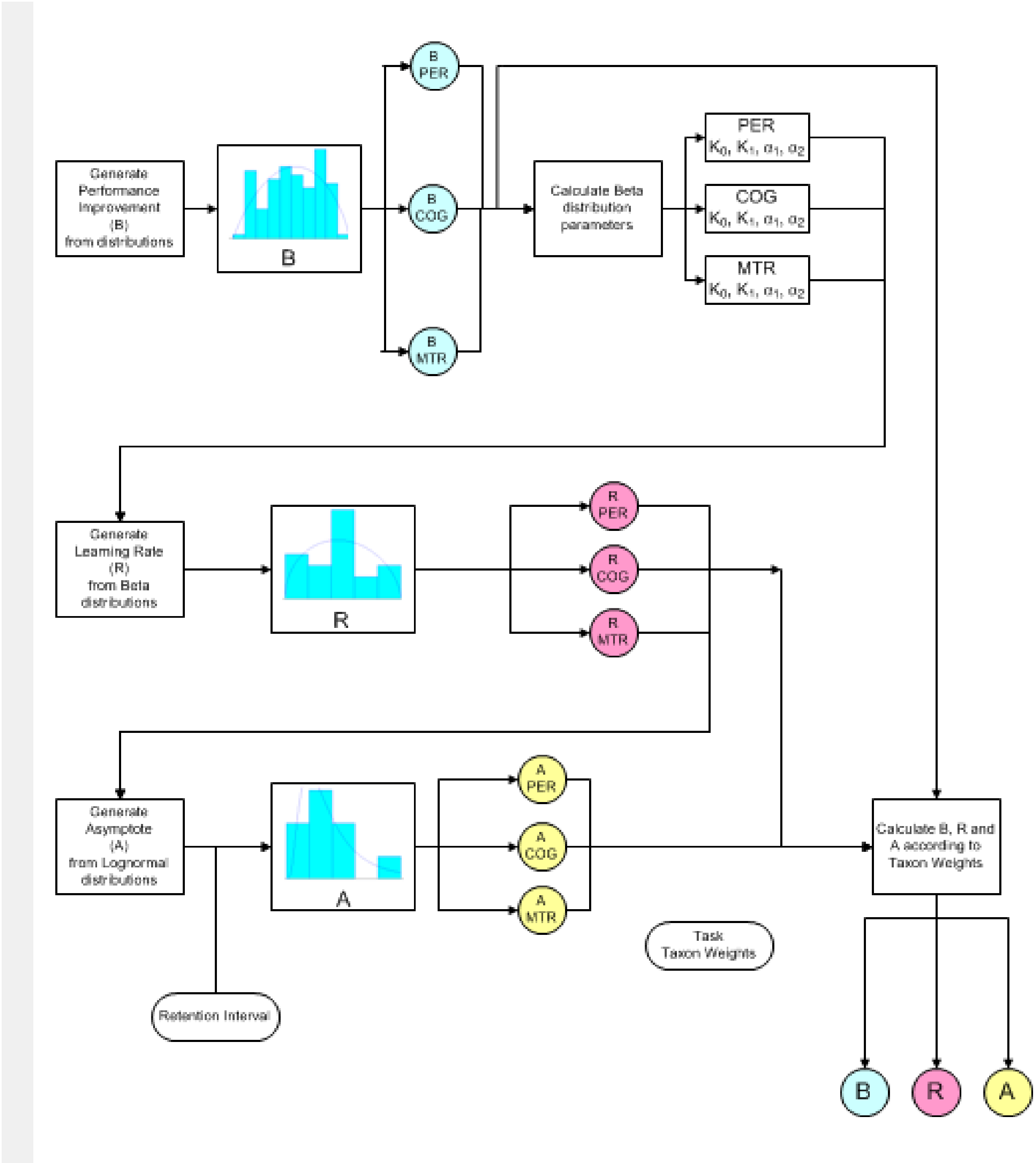


Figure 31. Flowchart of the core parameter modeling process.

With the three core parameters for a predicted learning curve now available, the power function equation

$$T = A + BN^{-R} \tag{4}$$

is used to generate performance indexes as shown in Figure 32.

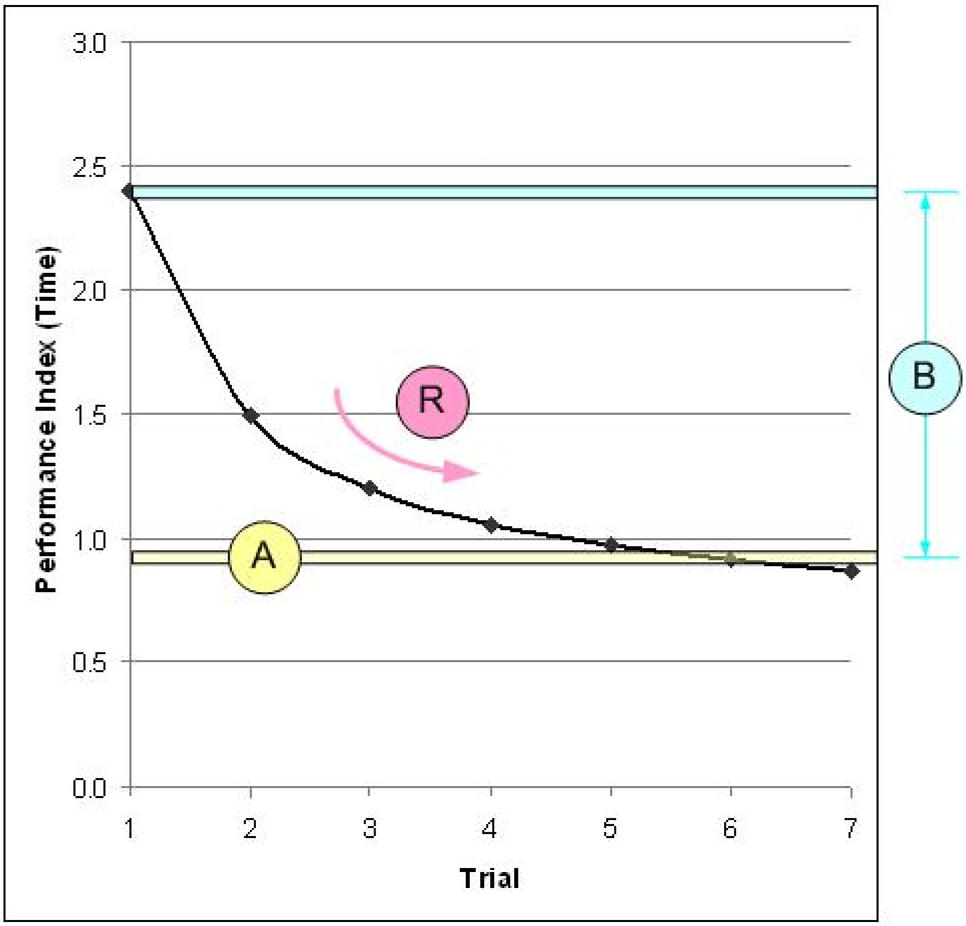


Figure 32. Learning curve generated by the core parameter modeling process.

The modeled data are plotted against actual data for comparison. Observations for the trials are averaged for the actual and model data with 95% confidence intervals to show the variance for each trial. Figures 33 through 35 show example plots for the low interactivity training strategy (TS =1), initial training condition (retention interval (RI) = 0) for each of the three isolated-taxon tasks. These plots are built by plotting the mean performance index for each trial along with that trial's 95% confidence interval.

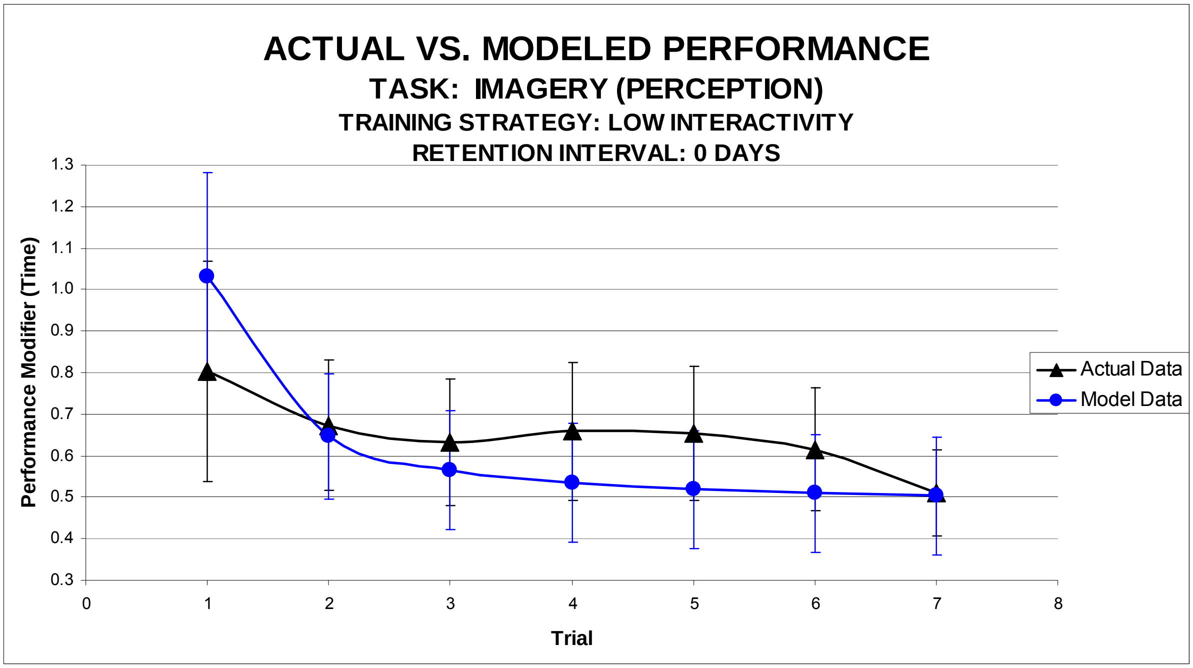


Figure 33. Plot of actual vs. fitted data for imagery task (TS=1, RI=0).

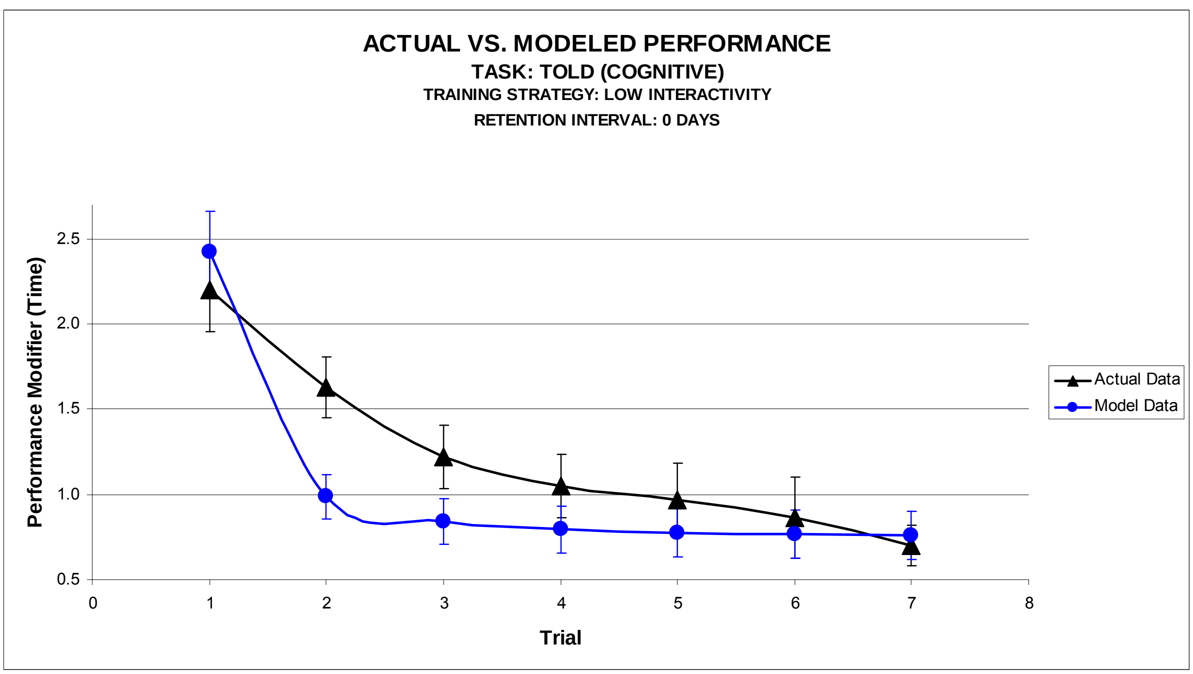


Figure 34. Plot of actual vs. fitted data for TOLD task (TS=1, RI=0).

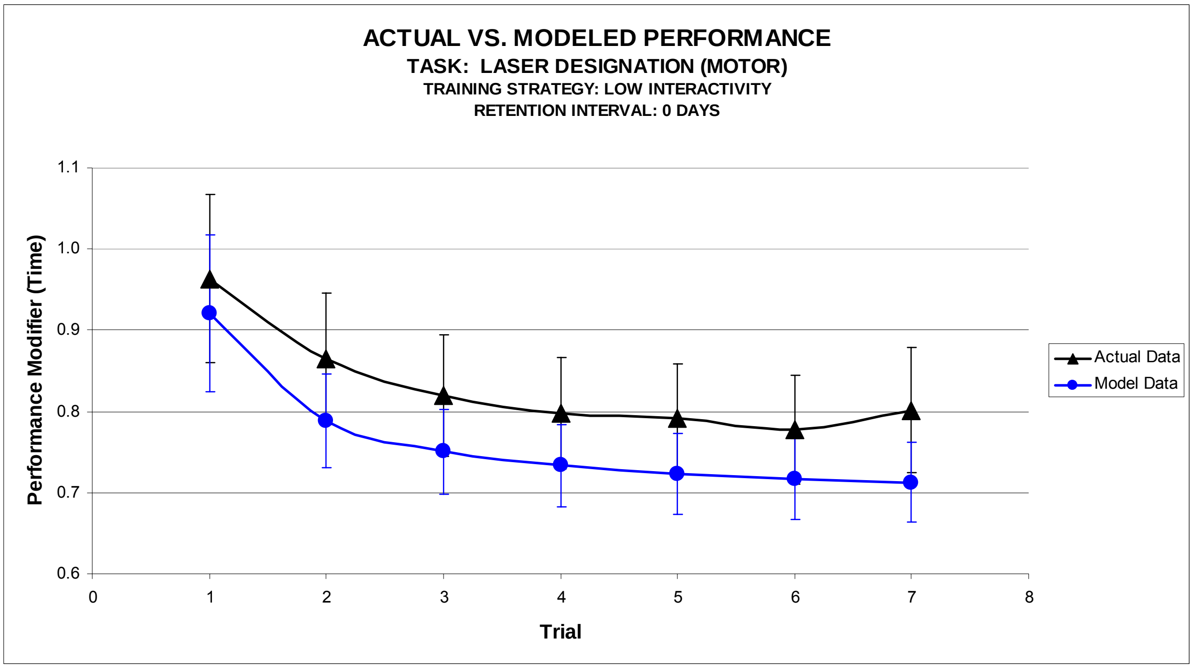


Figure 35. Plot of actual vs. fitted data for laser designation task (TS=1, RI=0).

6.2.2. Model Analysis 1

The results between the models for each of the tasks show varying levels of model accuracy. By visual inspection, the laser designation task shows a highly-correlated relationship between the actual and fitted data. Not only are values for the means of each of the trials very close, but the variances produced by the models also appear comparable.

The TOLD task model is very accurate for the first trial, but a steep descent to trial two indicates the model may be too sensitive at this point in the simulated training cycle. As trials continue, the predictions become more accurate.

The imagery task predictions appear to be more accurate that the TOLD prediction but less accurate than the laser designation model. While specific values for each trial mean deviates slightly, confidence intervals for the model significantly overlap those of the actual data suggesting the model would predict performance for twenty simulated participants, similar to the twenty participants who actually performed.

6.3. Model Validation 2

The second model validation serves to test the assertion that performance on new tasks can be predicted from models built from data collected in the performance of tasks that require certain human performance capabilities such as perception, cognition and motor skills. Actual data collected during the talk-on task were compared to a model for that task derived from the baseline tasks that isolate the three taxon categories.

6.3.1. Model Execution 2

The model for this prediction is similar to that described in Section 6.2.1. However, because the talk-on task shares characteristics with all three tasks, core parameters for each of the taxon categories were drawn for twenty simulated participants each. These values were then averaged using a weighting scheme that suggests the weight each taxon category impacts the talk-on task. In this test, it is assumed the talk-on task taxon weighting is 0.6, 0.2 and 0.2 for the perceptual (imagery), cognitive (TOLD), and motor skill (laser designation) taxon categories, respectively. These weighted values for B, R and A were then used to generate twenty simulated participants performing the talk-on task for comparison with real data. Figure 36 shows the results of one such test for the low interactivity (TS=1) training strategy, retention interval for the initial training period (RI=0).

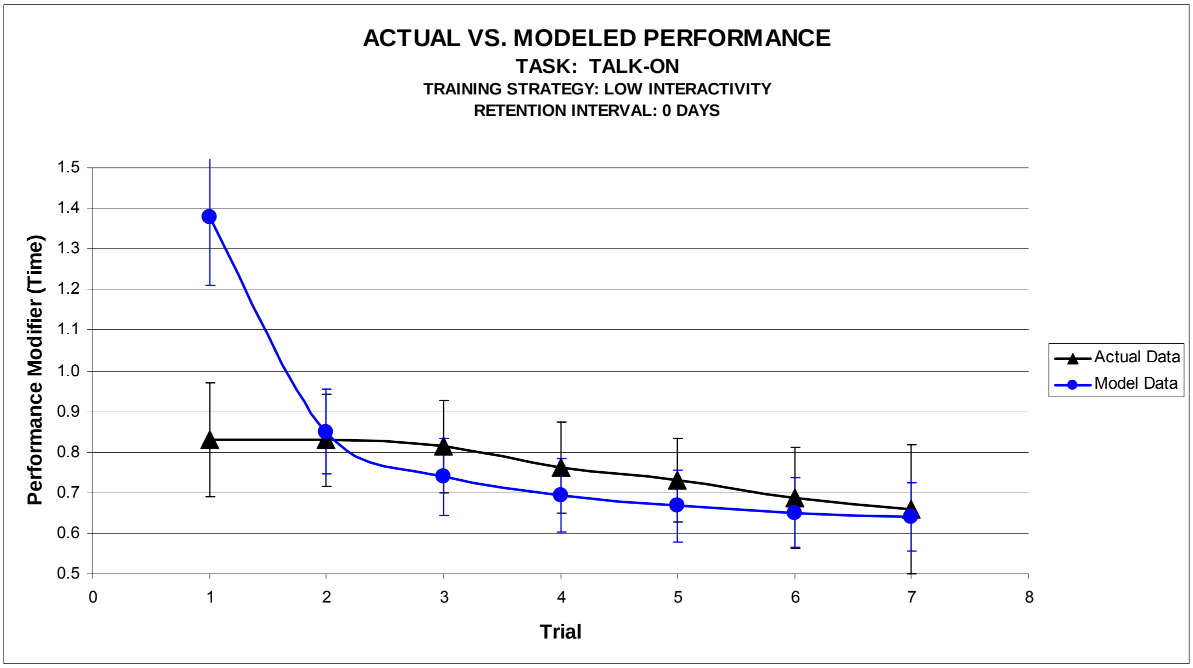


Figure 36. Plot of actual vs. fitted data for the talk-on task (TS=1, RI=0).

6.3.2. Model Analysis 2

The performance of the model is most similar to the imagery task. While trial number 1 shows a modeled mean significantly higher than the actual mean (and 95% confidence intervals that do not intersect), mean and variance on subsequent trials are similar. This finding suggests this modeling approach is a valid method to predict training outcomes on tasks for which there is no actual performance data.

See Appendices D through G for actual and fitted data comparisons for all tasks, training strategies and retention intervals.

7. Summary and Discussion

7.1. Summary

Determining and improving training effectiveness is necessary when considering the resources required for training, particularly in the context of organizations with enormous training concerns such as the military (Salas, et al., 1995). While the body of work attempting to mathematically characterize human learning is extensive, there have been very few attempts at applying well accepted modeling approaches, using empirical data, to mathematical functions of learning.

Overall, the results of this research suggest there is significant merit in this modeling approach to warrant continued work.

7.2. Benefits, limitations and future work

This experiment is unique in its size, scope and approach to the collection of learning data. Sixty participants completed two data collection sessions that contributed to a dataset of over 4000 data points for time performance. Similar to other experiments of this nature, logistical necessity demanded certain compromises, particularly where the amount of time to train tasks and perform them is required. However, with the assistance of Predator sensor operator subject matter experts, tasks were developed that were learnable within a short timeframe, yet relatable to real world tasks.

The approach taken in this research is unique in that the method used to create the models of learning is chosen specifically to support a stochastic simulation. For the modeler who needs to simulate a population of trainees learning tasks with definitive characteristics, models based on key parameters in the learning process produce a dataset useful for a variety of performance predictions.

Lane's (1987) examination of the learning literature to find data that would support performance prediction suffered from widely varying performance parameters. This research followed an approach similar to performance moderator functions already installed in IMPRINT, specifically (1) classifying tasks according to basic human skills, and (2) building performance modifiers for each skill type (Laughery & Corker, 1997). While the body of learning research is extensive, empirical data was historically collected from tasks that were either too simple or otherwise not relatable to a real-world task (Lane, 1987).

The tasks used in this research were selected so they could be learned within a relatively short time period, yet they were complex enough in their difficulty and range of taxons to make them comparable to a realistic operational task.

Like all studies, this research is not without its limitations. While every effort was made to make this experiment as realistic as possible, compromises were made in the interest of logistical efficiency. Relating laboratory measures to actual field data has been problematic for years and is difficult to accomplish without some inaccuracies (Laughery & Corker, 1997). Foremost of the differences between the data collected and real-world operations are (1) the inherent abilities and motivations of the participants involved, (2) the compressed training schedule that may induce different learning rates when compared with actual training conditions, and (3) the ability to properly characterize tasks by a meaningful task taxonomy.

There are several potential areas for future work in the domain. First, the issue of dynamically changing taxons during the learning process should be addressed. This research applied taxon weights based on what primary human performance channel is undergoing "reconfiguration" as a result of the training process. Modeling the changes between taxon weights at the beginning of training with those at the end of training would provide insight into not only the training process, but will also shed light on complex interactions between the channels of human performance. Continuing the discussion on taxon weightings is also important – because there is such variability between individuals on overall performance as seen by performance improvement, learning rates and asymptote, it seems logical there would be differences between individuals regarding difficulty of a task and what human performance channels are really being configured (i.e., is it reasonable to assume that everyone learning a task is learning about the same components of the task at the same rates).

7.3. Contributions of this research

This research combined well-accepted theories and techniques of two distinct, yet relatable fields of modeling and simulation and learning theory. While the main goal of this research was to solve a problem of one community using the process techniques of another, there are additional contributions worthy of discussion. These contributions are categorized according to their practical or theoretical implications.

7.3.1. Practical Implications

Though there is still vast potential for improvement, the modeling approach described has promise in a variety of modeling and analysis situations. The foremost benefit is in the ability to predict task performance as a function of variables typically manipulated in training system design, particularly the selection of training strategy that best instructs a given task. Given a large enough data set with which to build a more advanced model, this process provides the flexibility to make predictions significantly beyond the tasks used to construct the model.

While the primary purpose of the model is to predict performance outcomes to evaluate the effectiveness of training methods, it also provides a framework to engineer the makeup of a task. Consider a task that is only suitably taught using a single training strategy; what if the performance prediction for that task shows that most people will require too many iterations, or early iterations simply take too much time and therefore too many resources? The model could be used to optimize the performance time by varying taxon weights to a level that makes the training approach under study feasible. The challenge to the training designer then would turn to re-engineering the task by adjusting the proportions of human performance channels required to accomplish it (e.g., reduce cognitive load by providing performance aids).

7.3.2. Theoretical Implications

Beyond the practical applications of this modeling approach are theoretical underpinnings that contribute to this knowledge area. While many variables in the learning process were explicitly addressed such as training strategy, retention interval and task type, there were variables that are either difficult to quantify or correlate to specific performance outcomes. Among these factors are aptitude and motivation. Assuming that the population from which performance data are collected to build a model using this process is large enough, it seems logical to assume that these difficult-to-model variables are implicitly characterized in the model.

7.4. Conclusions

Advances in military weapons systems provide an ever increasing impact on the training community charged with developing an effective human operator to become an important element of those systems. According to Salas and Cannon-Bowers (2001), as pressure to ensure maximum training effectiveness grows "training practitioners will do well to employ sound principles, guidelines, specifications, and lessons learned from the literature, rather than relying on a trial-and error approach . . . a new era of training has begun – one in which a truly reciprocal relationship between training research and practice will be realized".

This study extended the body of knowledge on modeling human performance during training by combining accepted practices and theories from two separate bodies of research – learning theory and modeling and simulation. This research represents an important first step in the field of training effects modeling and shows promise to effectively predict the resources required to train the human operator.

APPENDIX A: DISTRIBUTION MODELS FOR CORE PARAMETER PERFORMANCE IMPROVEMENT (B)

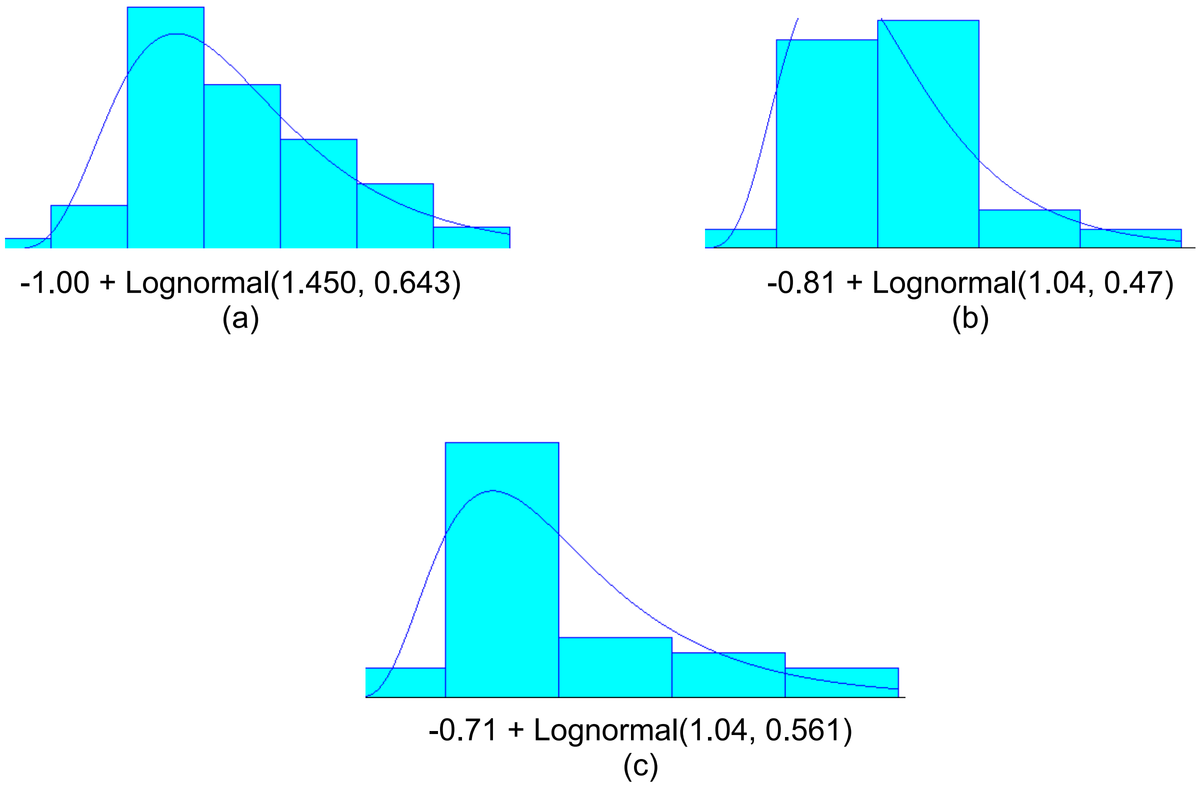
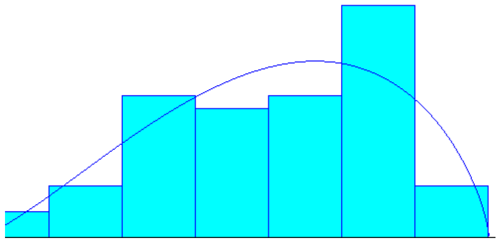
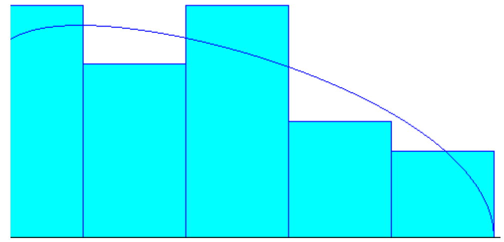


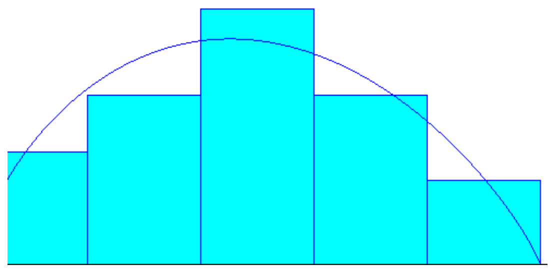
Figure 37. Distribution for performance improvement (B) for imagery task for retention intervals of (a) 0 days, (b) 30 days, and (c) 60 days.



0.06 + 2.77*Beta(2.36, 1.7)
(a)



0.21 + 2.28*Beta(1.12, 1.51)
(b)



0.19 + 2.52*Beta(1.69, 1.84)
(c)

Figure 38. Distribution for performance improvement (B) for TOLD task for retention intervals of (a) 0 days, (b) 30 days, and (c) 60 days.

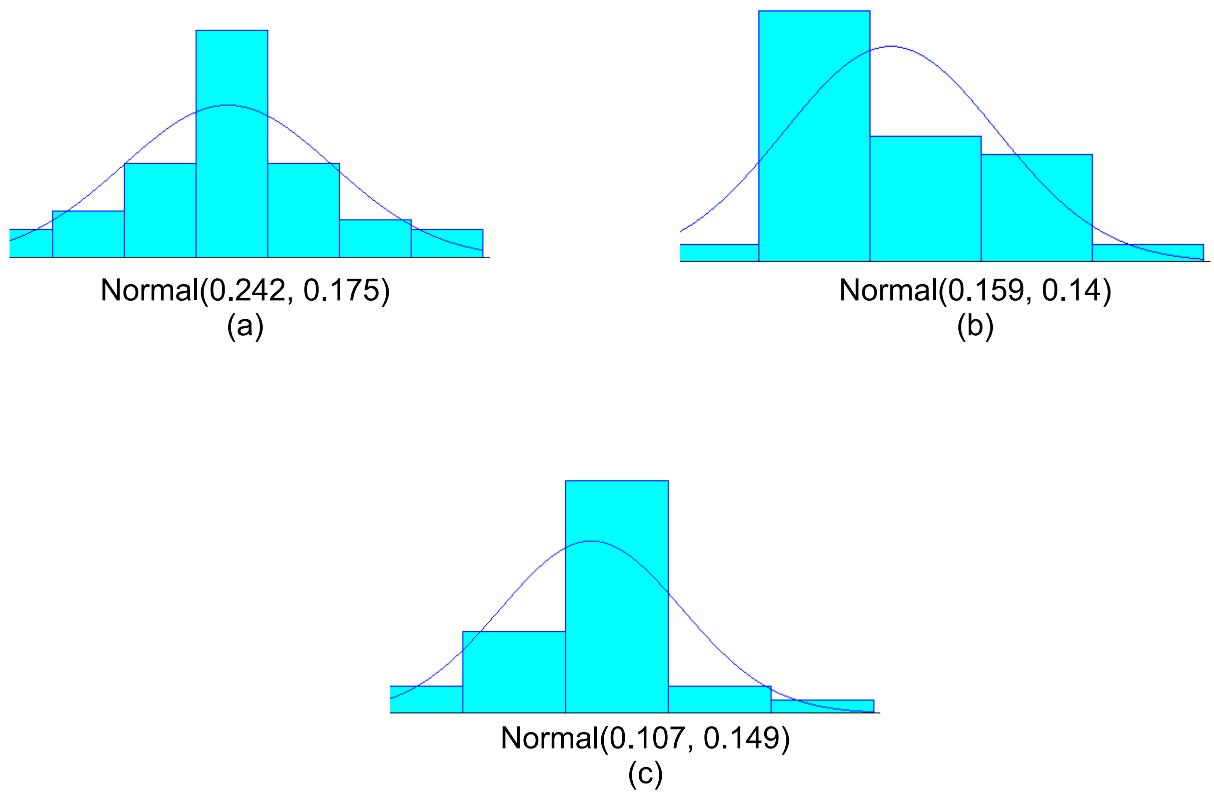


Figure 39. Distribution for performance improvement (B) for laser designation task for retention intervals of (a) 0 days, (b) 30 days, and (c) 60 days.

APPENDIX B: DISTRIBUTION MODELS FOR CORE PARAMETER LEARNING RATE (R)

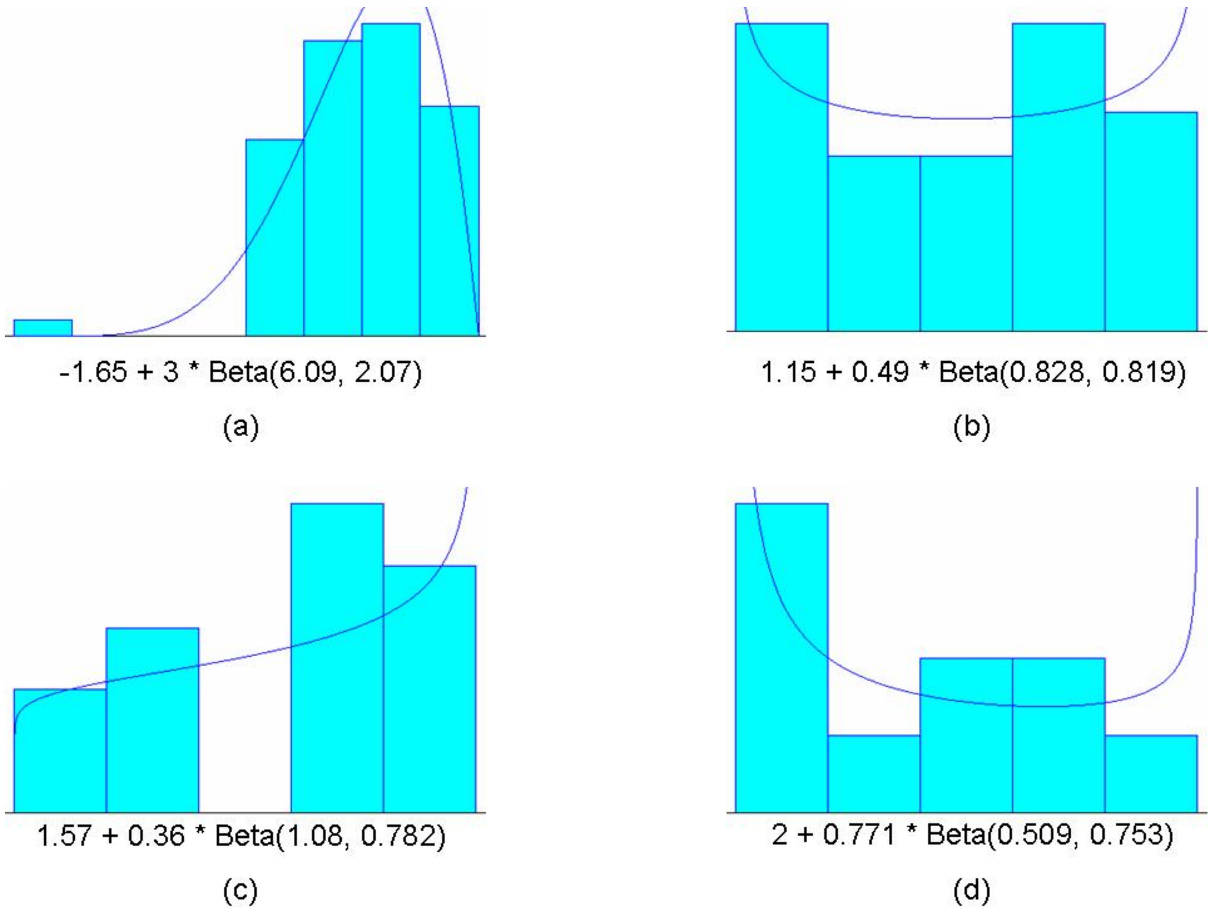


Figure 40. Distributions for learning rate (R) for B values of (a) 0.0, (b) 0.5, (c) 1.0, and (d) 1.5 for imagery task.

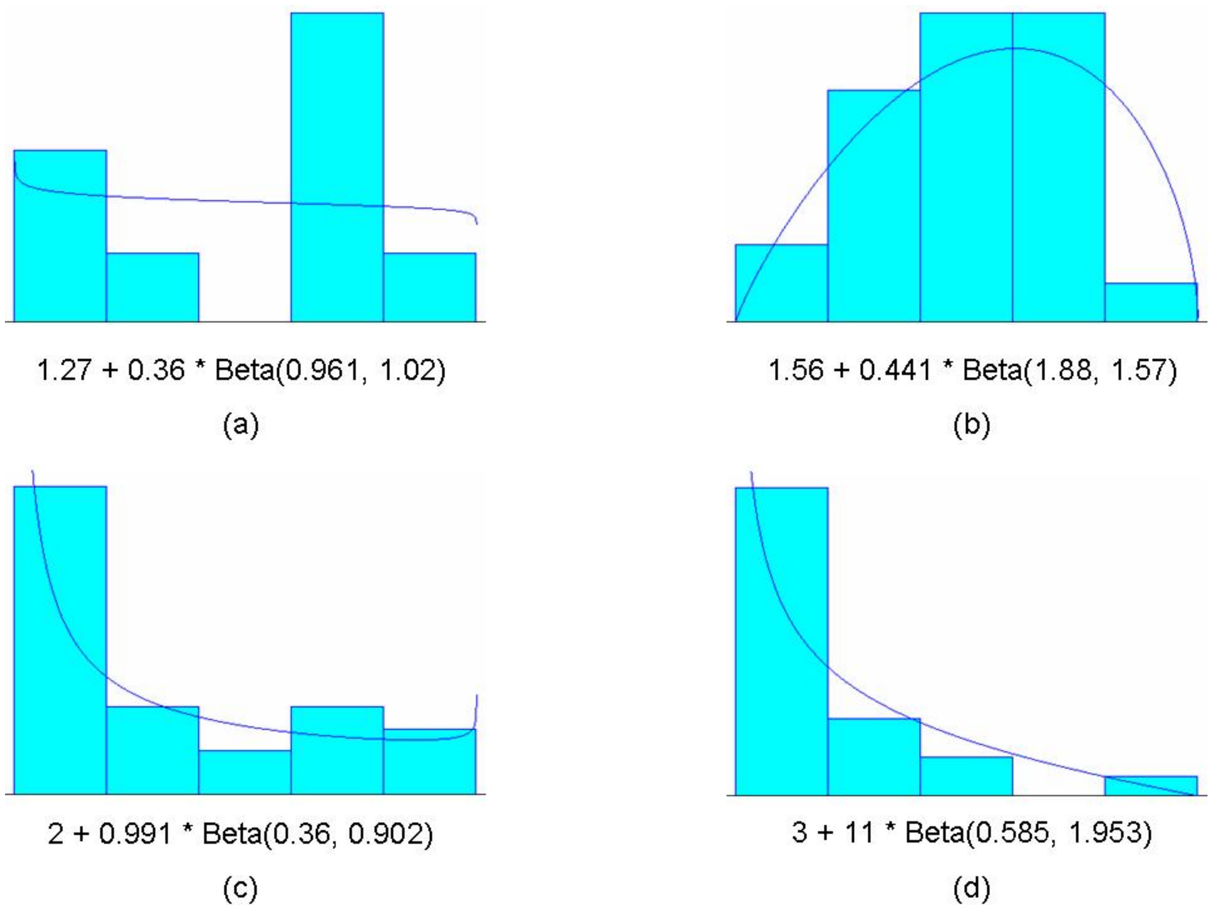


Figure 41. Distributions for learning rate (R) for B values of (a) 0.5, (b) 1.0, (c) 1.5, and (d) 2.0 for TOLD task.

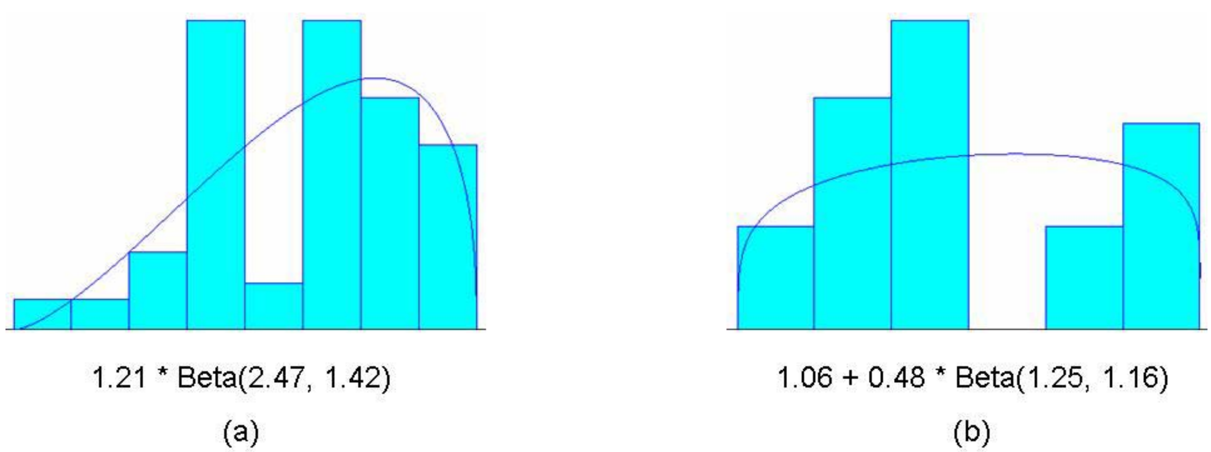


Figure 42. Distributions for learning rate (R) for B values of (a) 0.0 and (b) 0.5 for laser designation task.

APPENDIX C: DISTRIBUTION MODELS FOR CORE PARAMETER ASYMPTOTE (A)

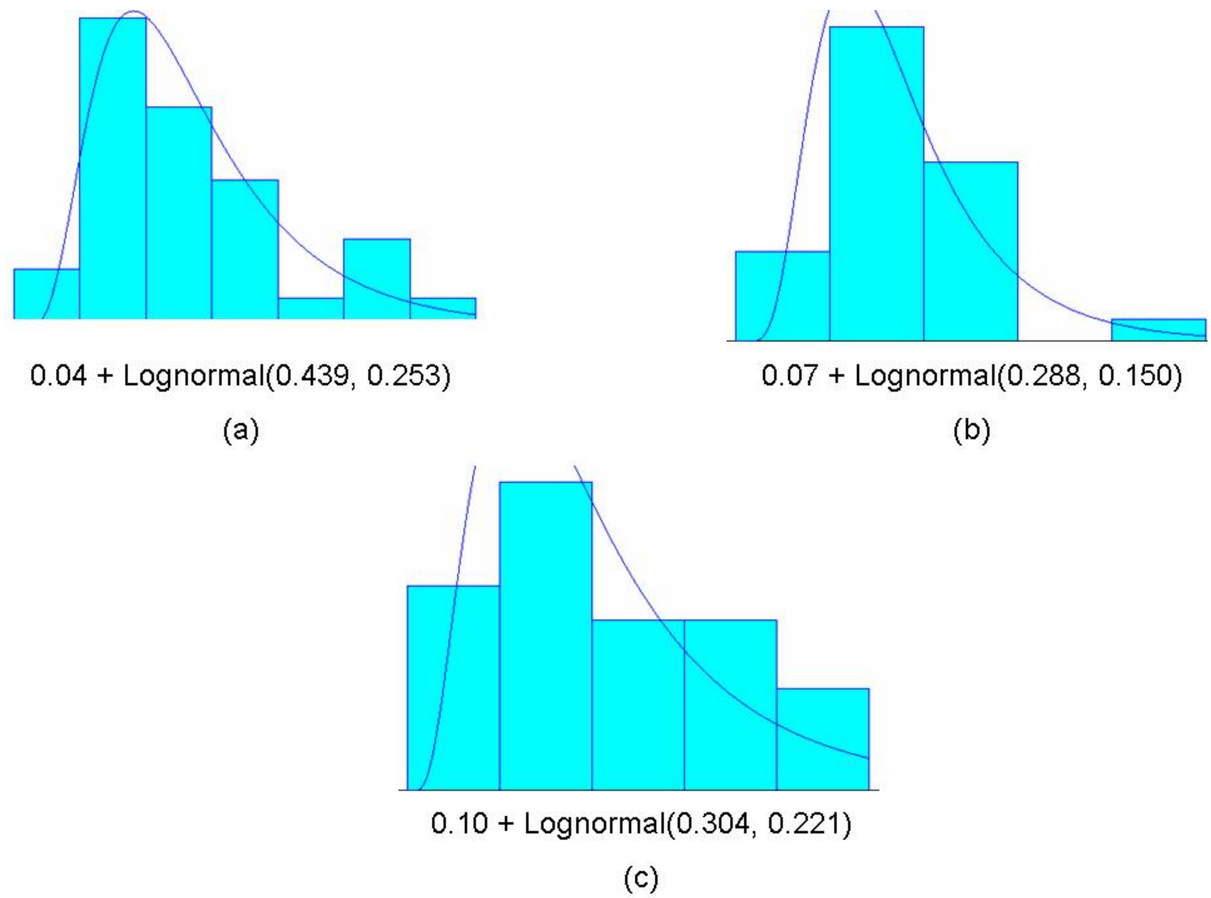
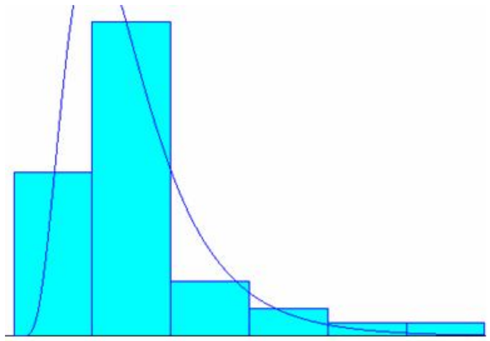
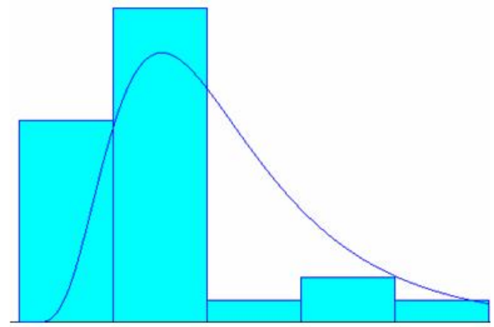


Figure 43. Distributions for asymptote (A) for retention interval values of (a) 0, (b) 30 and (c) 60 days for imagery task.



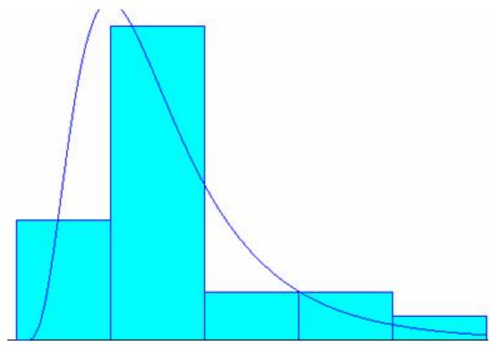
0.12 + Lognormal(0.475, 0.263)

(a)



0.19 + Lognormal(0.309, 0.176)

(b)



0.19 + Lognormal(0.446, 0.274)

(c)

Figure 44. Distributions for asymptote (A) for retention interval values of (a) 0, (b) 30 and (c) 60 days for TOLD task.

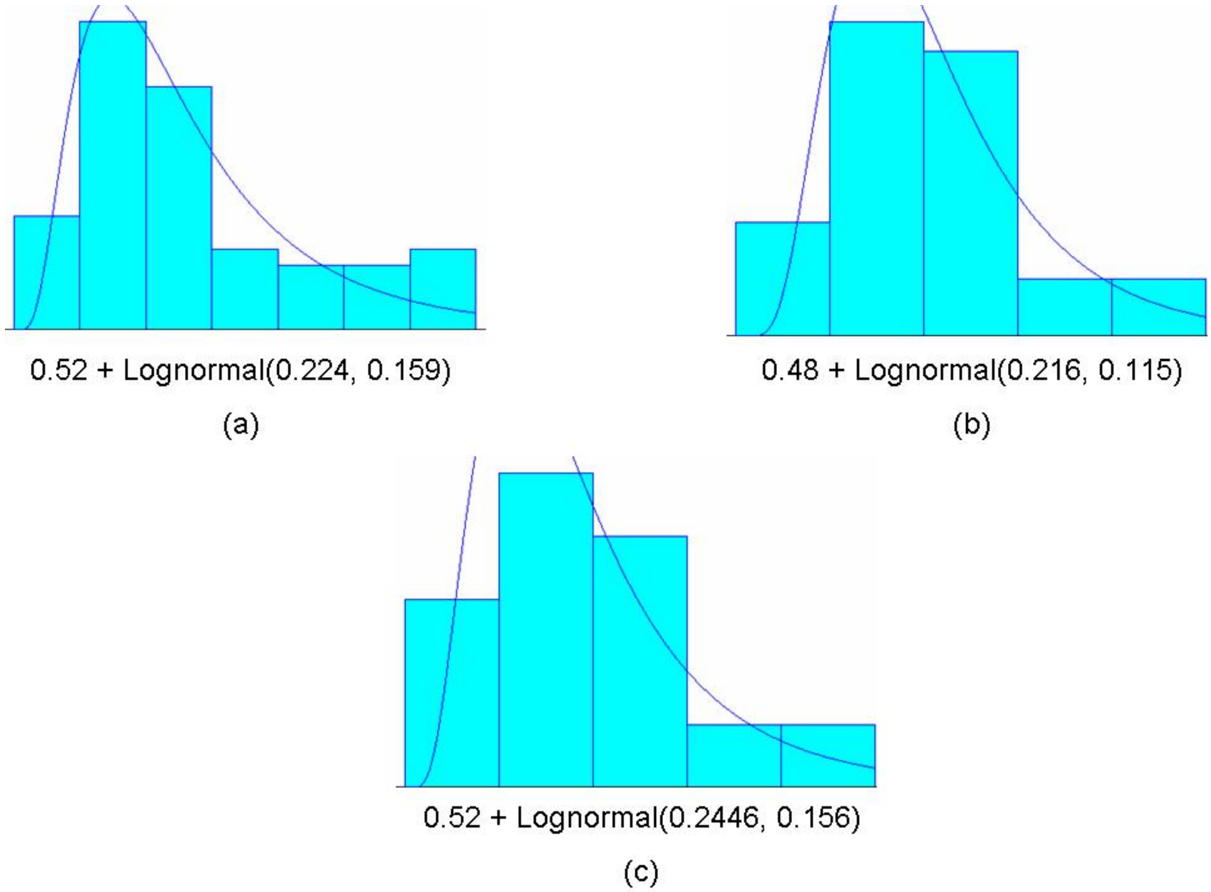
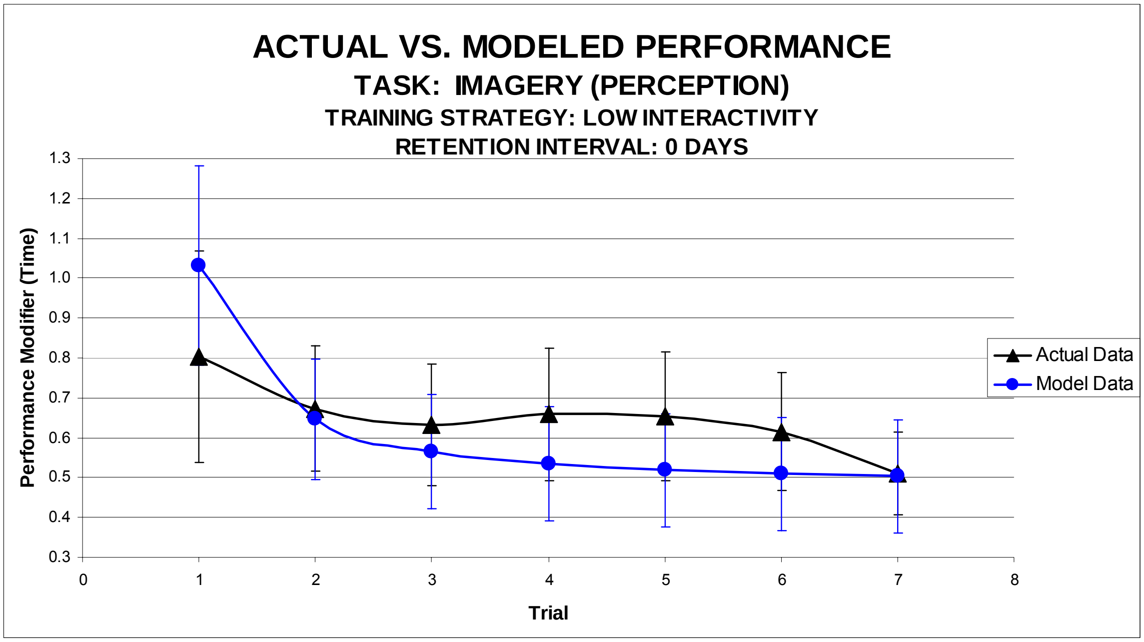
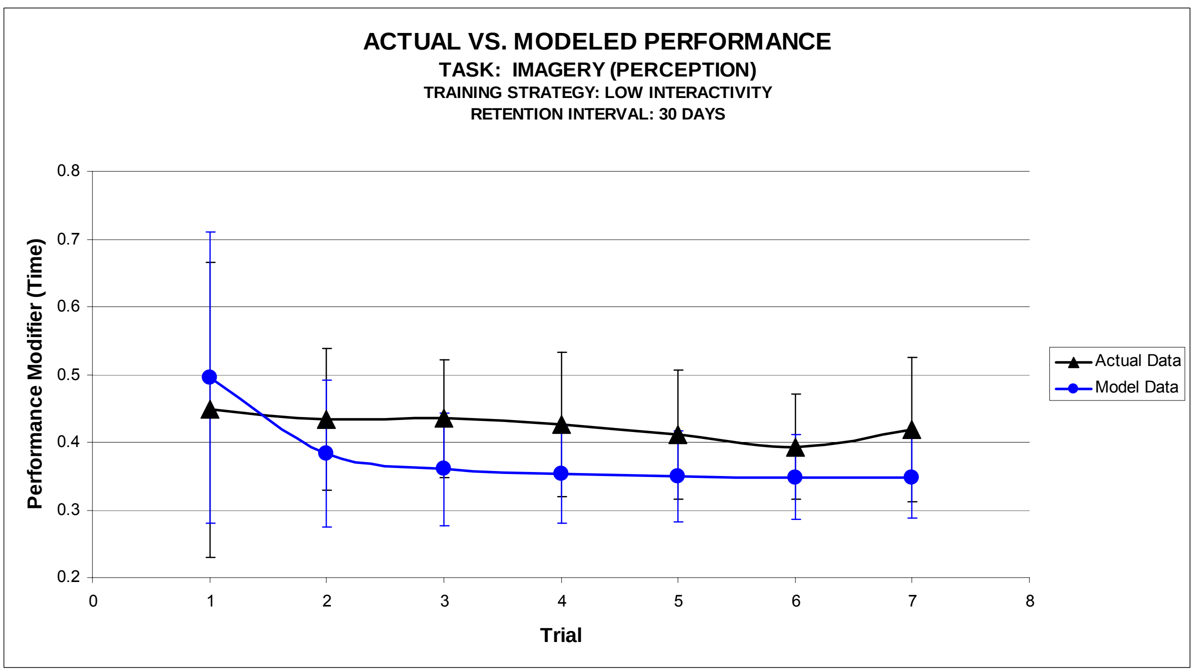
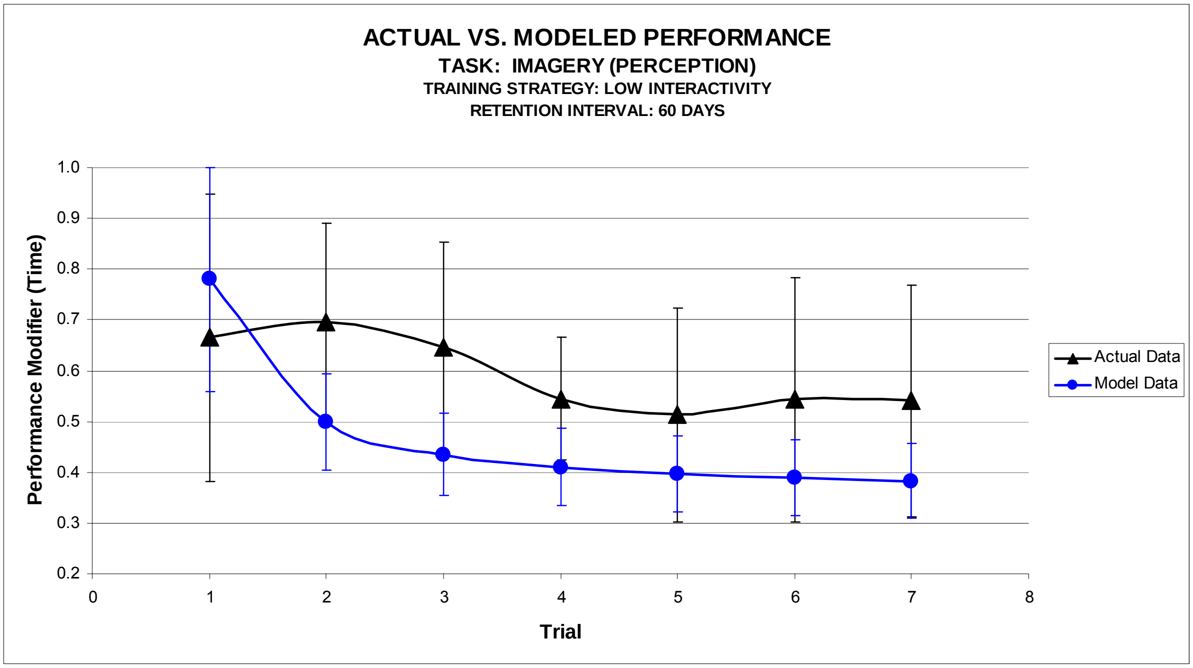


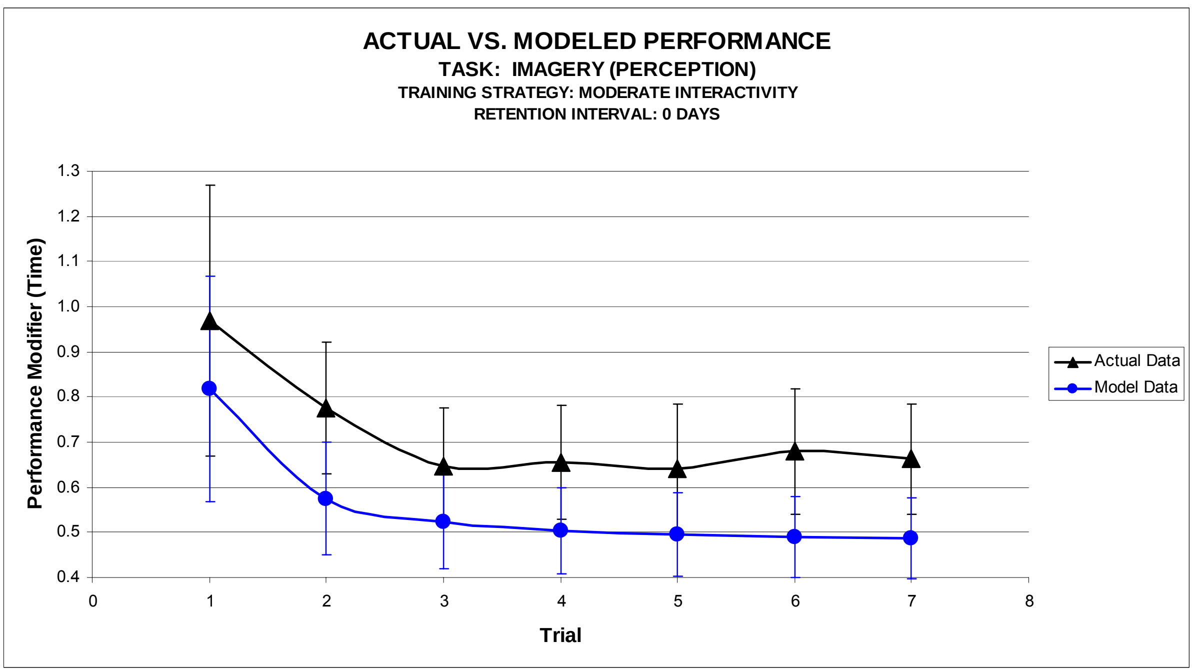
Figure 45. Distributions for asymptote (A) for retention interval values of (a) 0, (b) 30 and (c) 60 days for laser designation task.

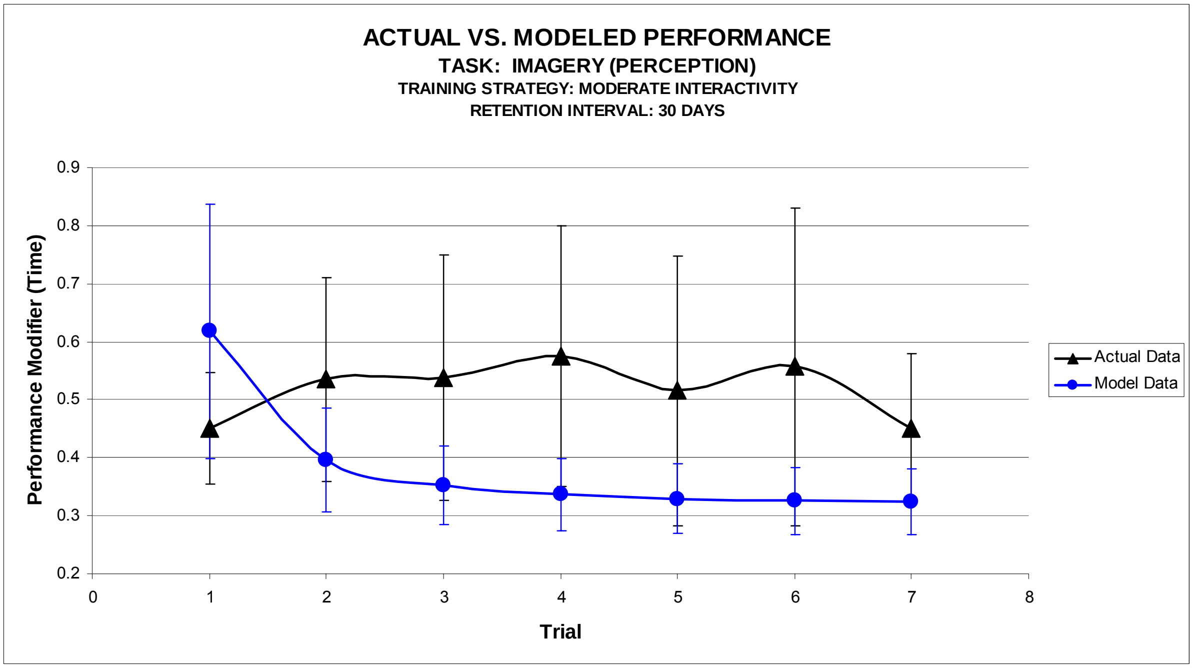
APPENDIX D: ACTUAL VS. FITTED DATA FOR PERCEPTION TASK (IMAGERY)

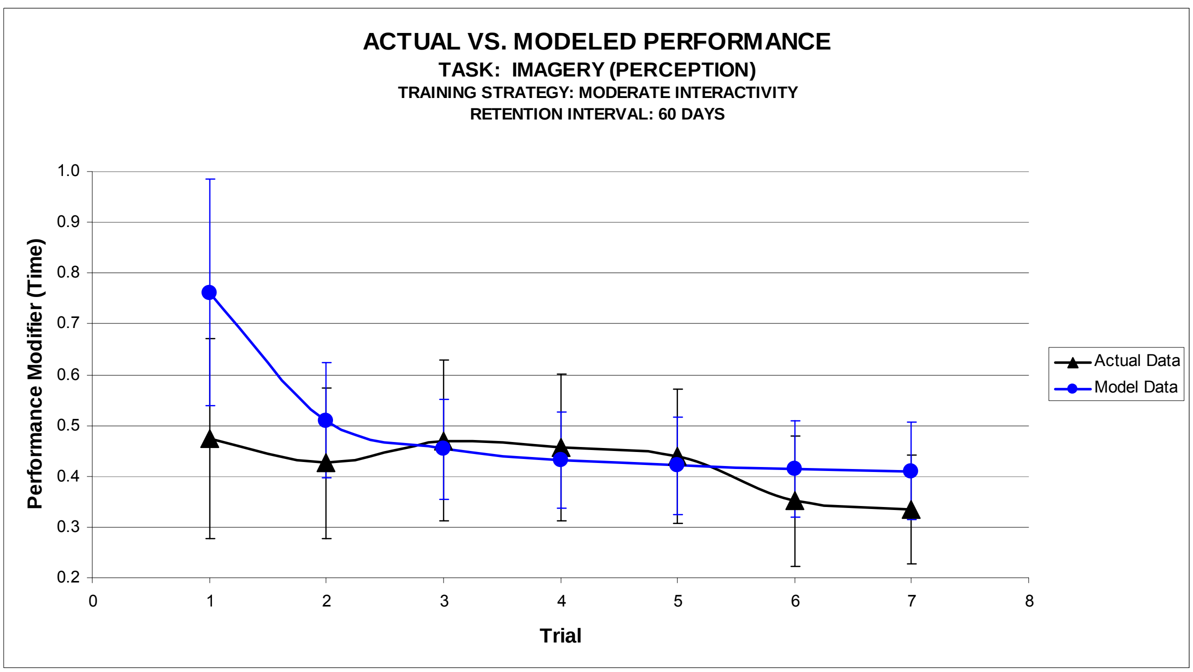


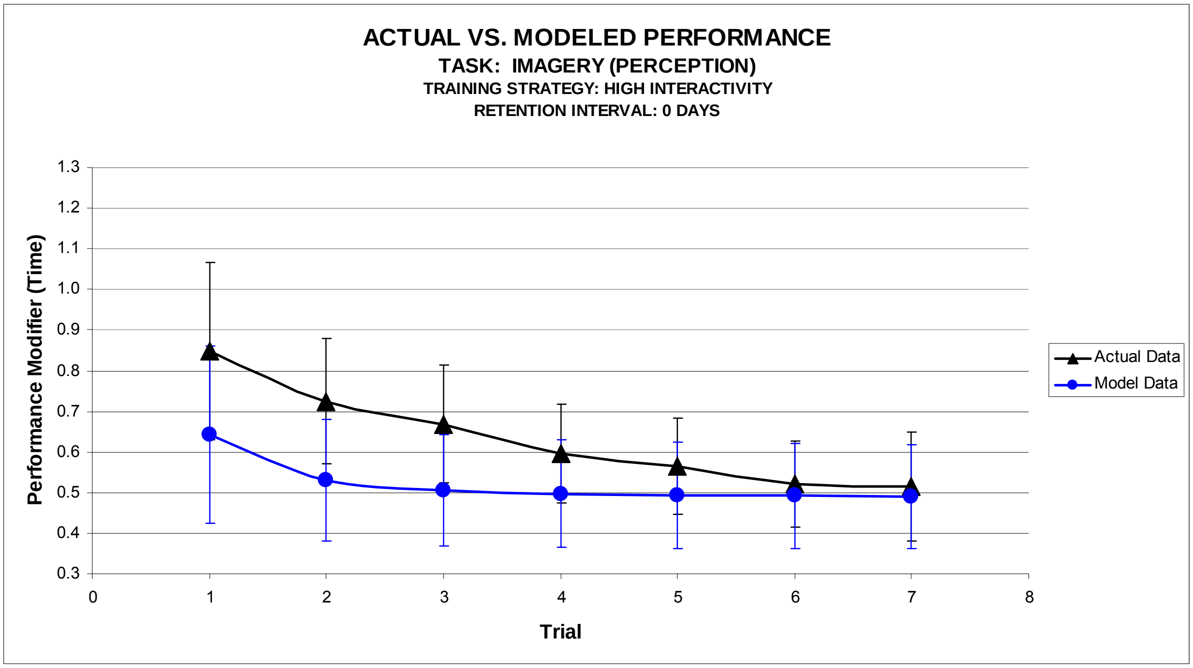


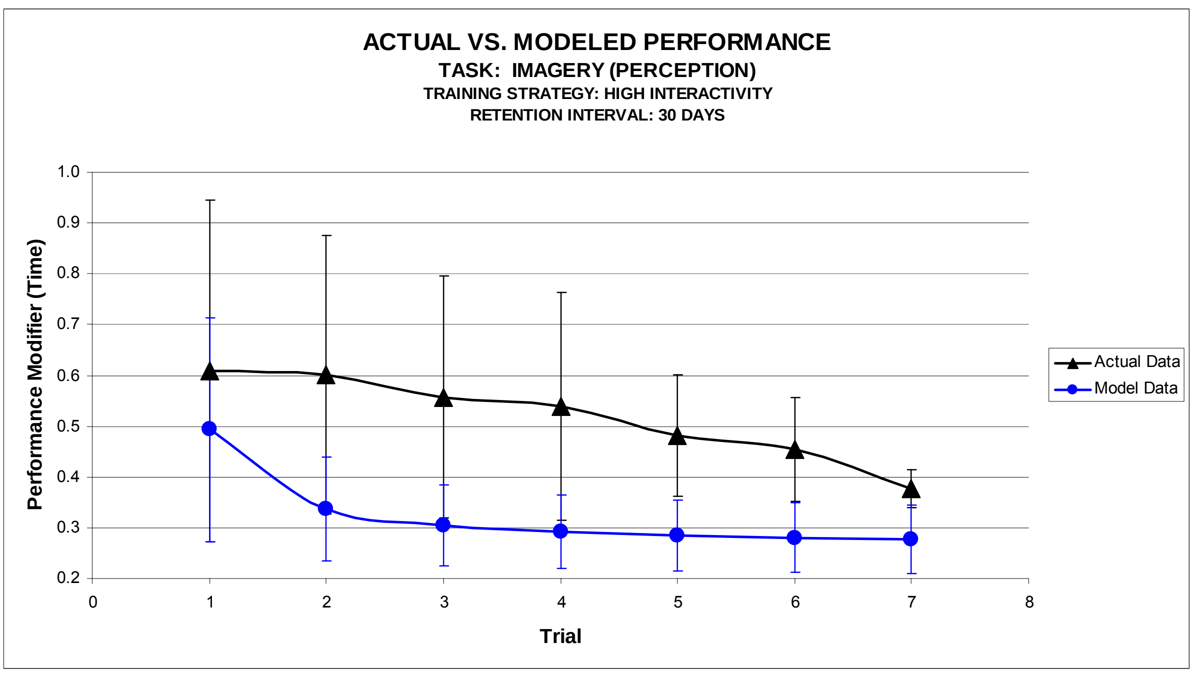

ACTUAL VS. MODELED PERFORMANCE
TASK: IMAGERY (PERCEPTION)
TRAINING STRATEGY: LOW INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

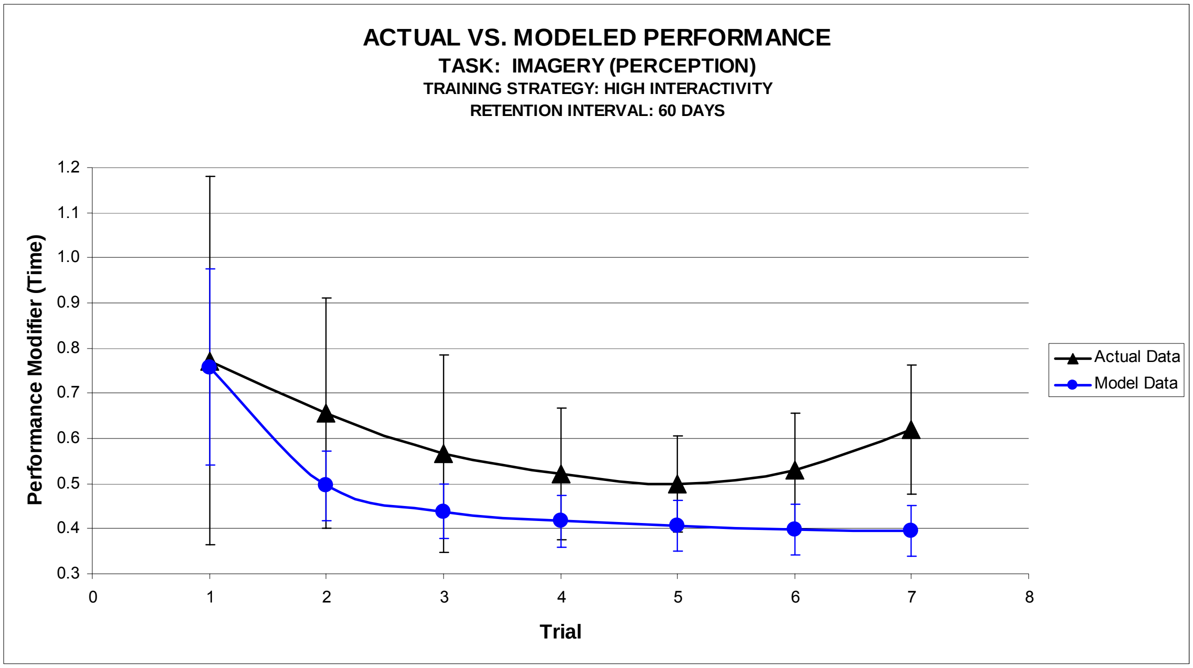

ACTUAL VS. MODELED PERFORMANCE
TASK: IMAGERY (PERCEPTION)
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 0 DAYS
Performance Modifier (Time)
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data


ACTUAL VS. MODELED PERFORMANCE
TASK: IMAGERY (PERCEPTION)
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 30 DAYS
Performance Modifier (Time)
Trial
Actual Data
Model Data

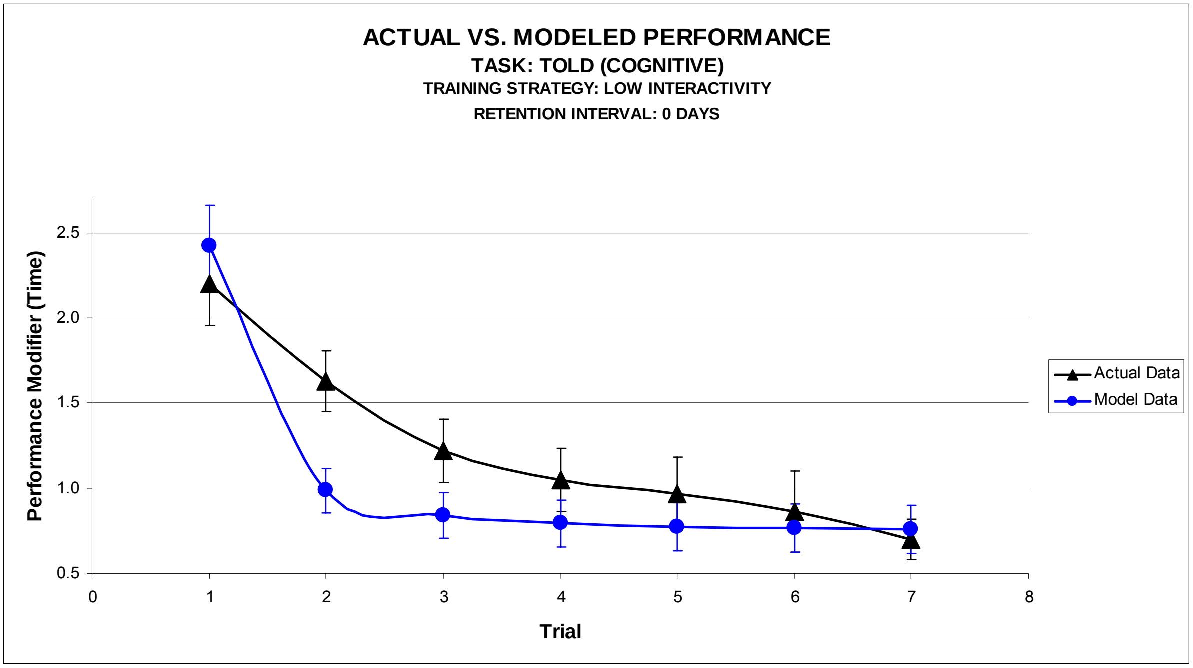

ACTUAL VS. MODELED PERFORMANCE
TASK: IMAGERY (PERCEPTION)
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
Trial
Actual Data
Model Data

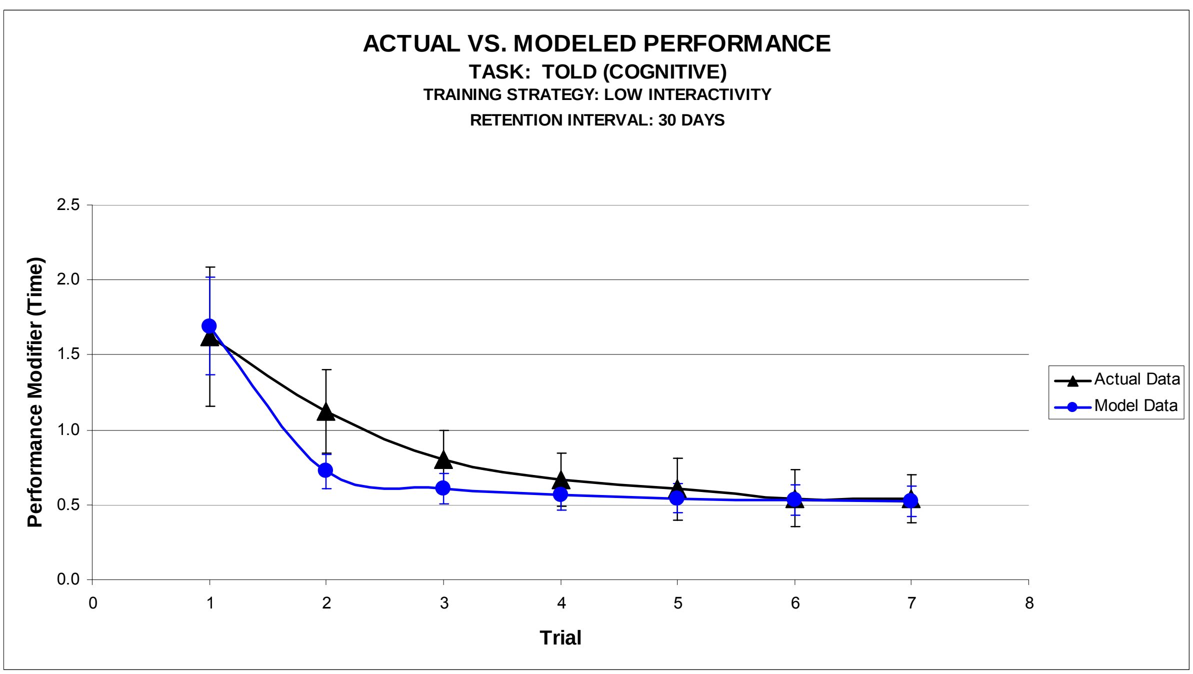

ACTUAL VS. MODELED PERFORMANCE
TASK: IMAGERY (PERCEPTION)
TRAINING STRATEGY: HIGH INTERACTIVITY
RETENTION INTERVAL: 0 DAYS
Performance Modifier (Time)
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

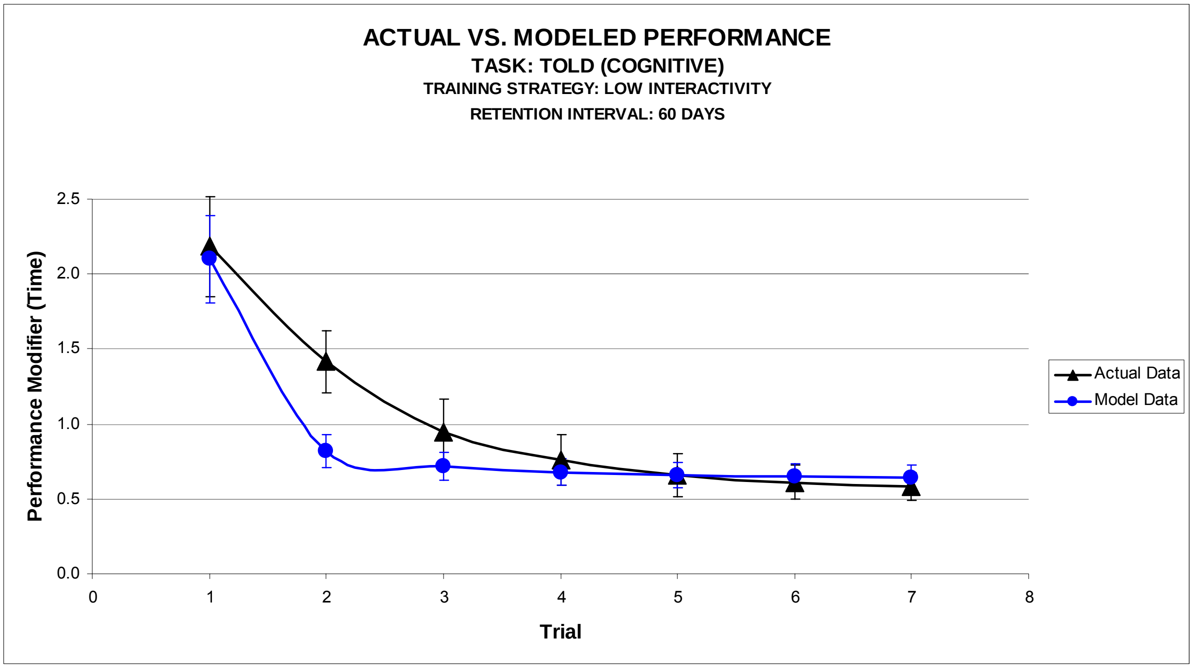

ACTUAL VS. MODELED PERFORMANCE
TASK: IMAGERY (PERCEPTION)
TRAINING STRATEGY: HIGH INTERACTIVITY
RETENTION INTERVAL: 30 DAYS
Performance Modifier (Time)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

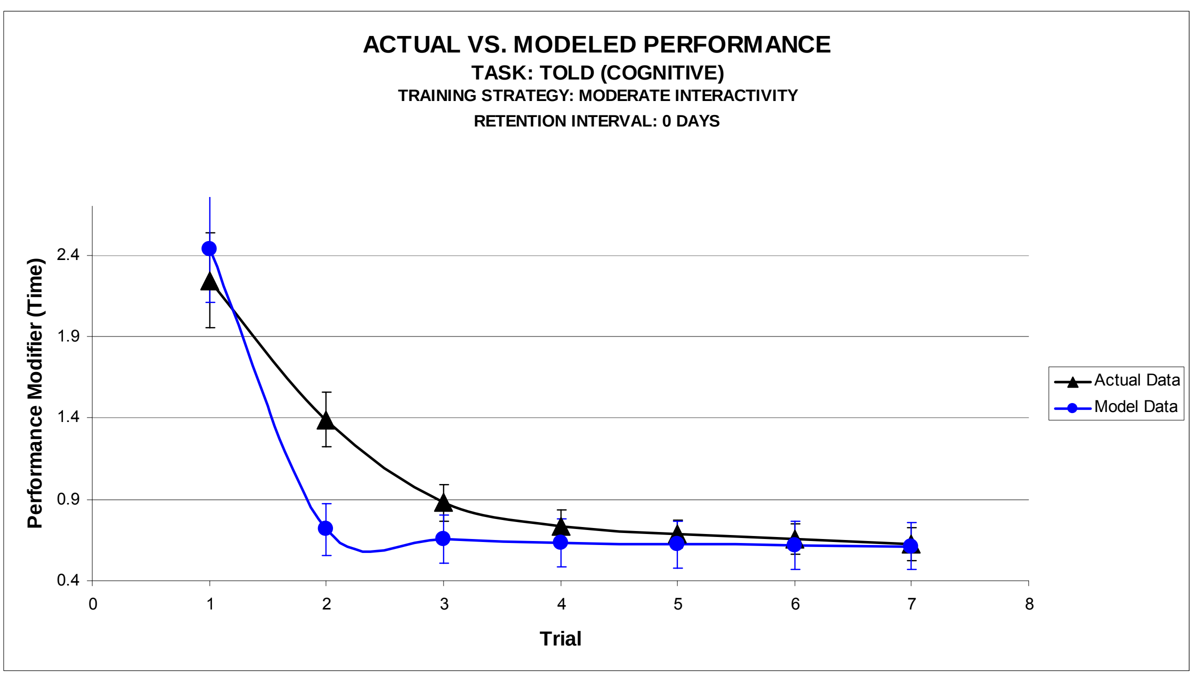

ACTUAL VS. MODELED PERFORMANCE
TASK: IMAGERY (PERCEPTION)
TRAINING STRATEGY: HIGH INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

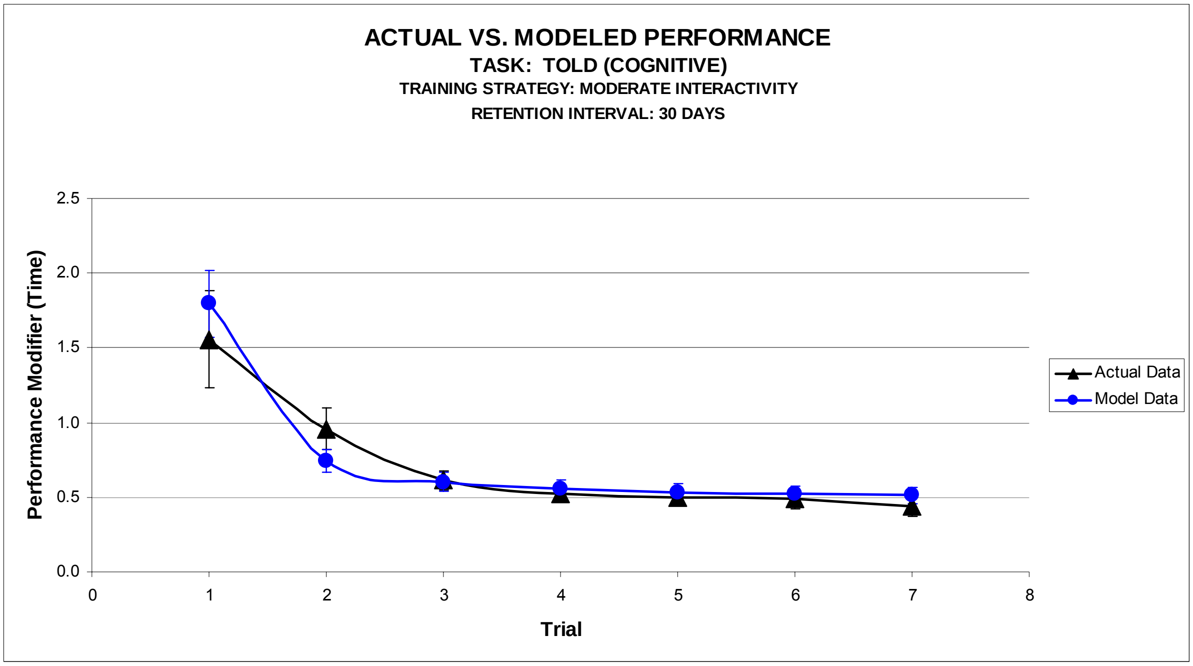
APPENDIX E: ACTUAL VS. FITTED DATA FOR COGNITIVE TASK (TOLD)

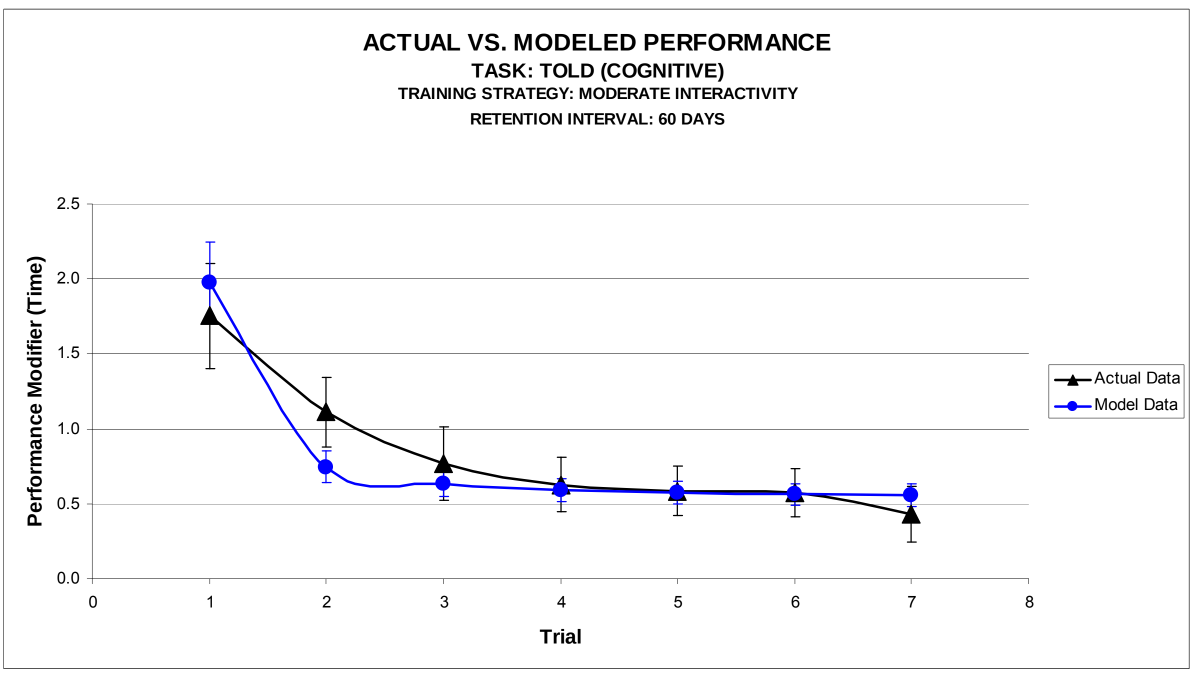


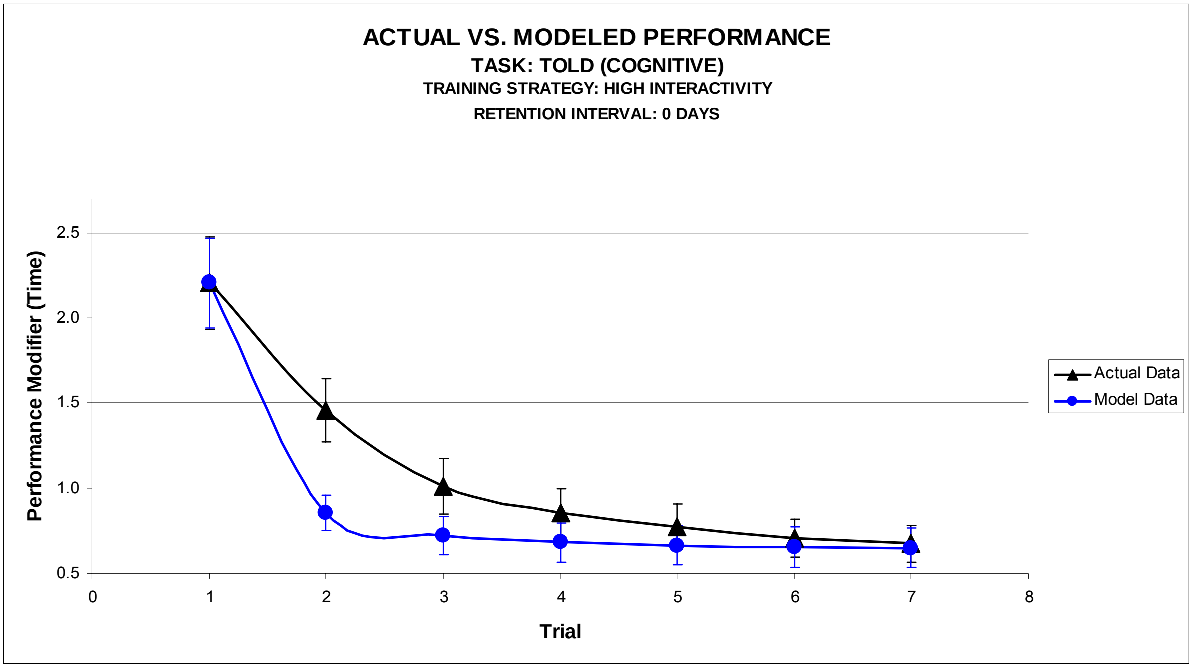



ACTUAL VS. MODELED PERFORMANCE
TASK: TOLD (COGNITIVE)
TRAINING STRATEGY: LOW INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

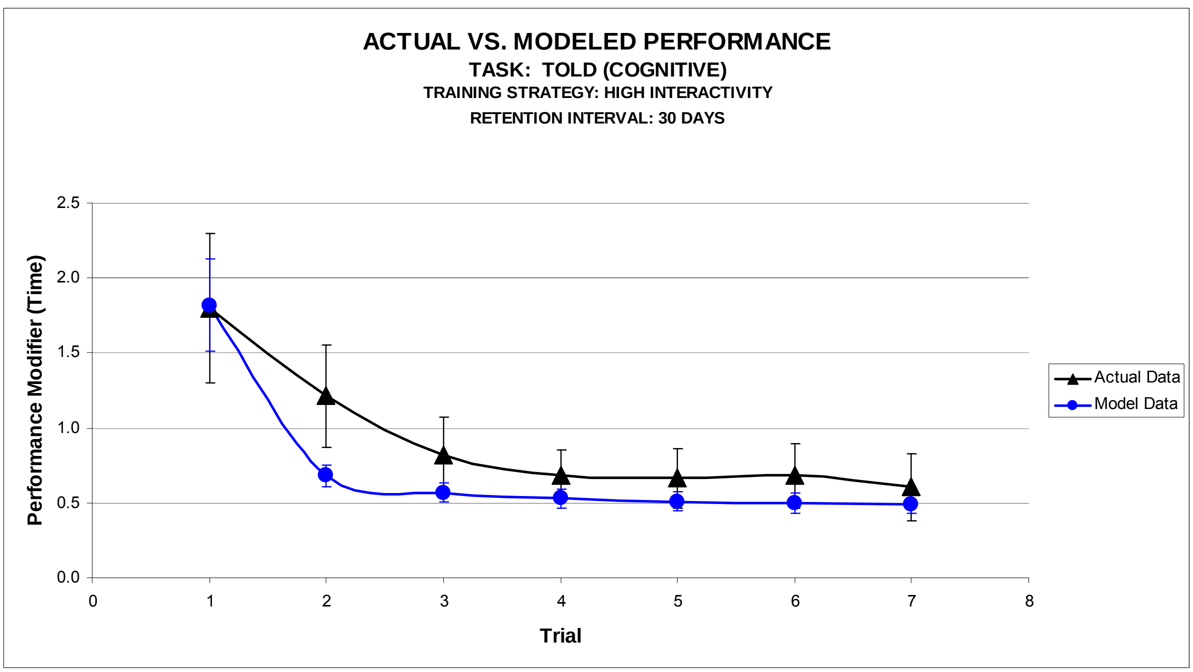

ACTUAL VS. MODELED PERFORMANCE
TASK: TOLD (COGNITIVE)
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 0 DAYS
Performance Modifier (Time)
2.4
1.9
1.4
0.9
0.4
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data


ACTUAL VS. MODELED PERFORMANCE
TASK: TOLD (COGNITIVE)
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 30 DAYS
Performance Modifier (Time)
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

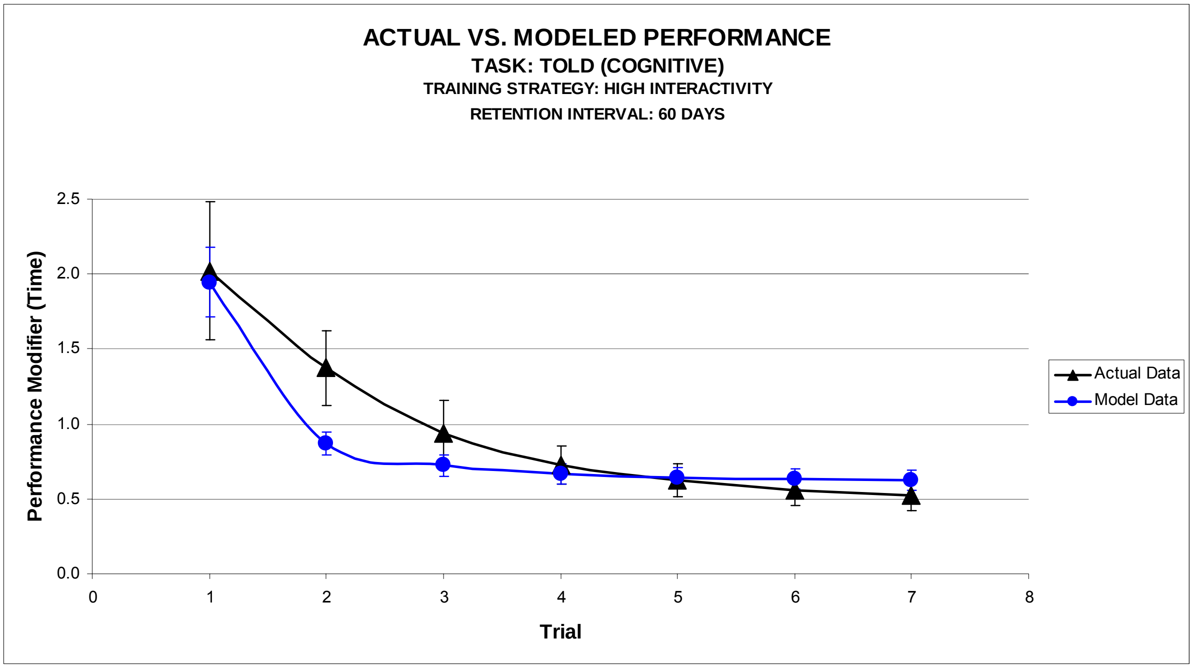

ACTUAL VS. MODELED PERFORMANCE
TASK: TOLD (COGNITIVE)
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data



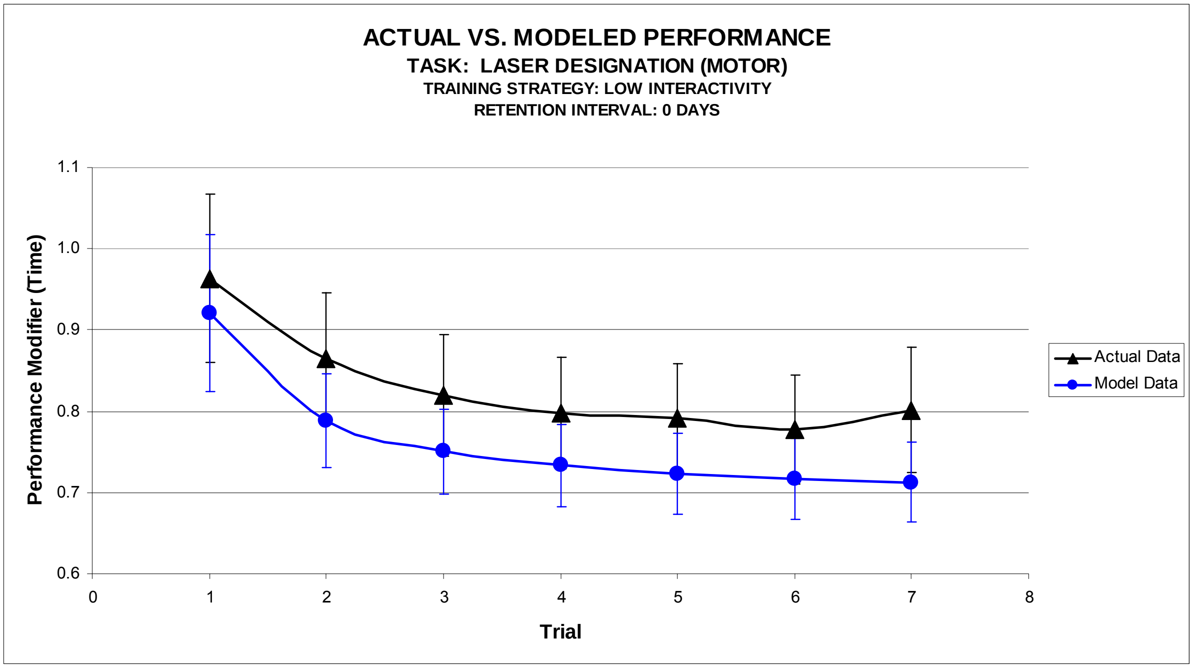
ACTUAL VS. MODELED PERFORMANCE
TASK: TOLD (COGNITIVE)
TRAINING STRATEGY: HIGH INTERACTIVITY
RETENTION INTERVAL: 0 DAYS
Performance Modifier (Time)
2.5
2.0
1.5
1.0
0.5
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

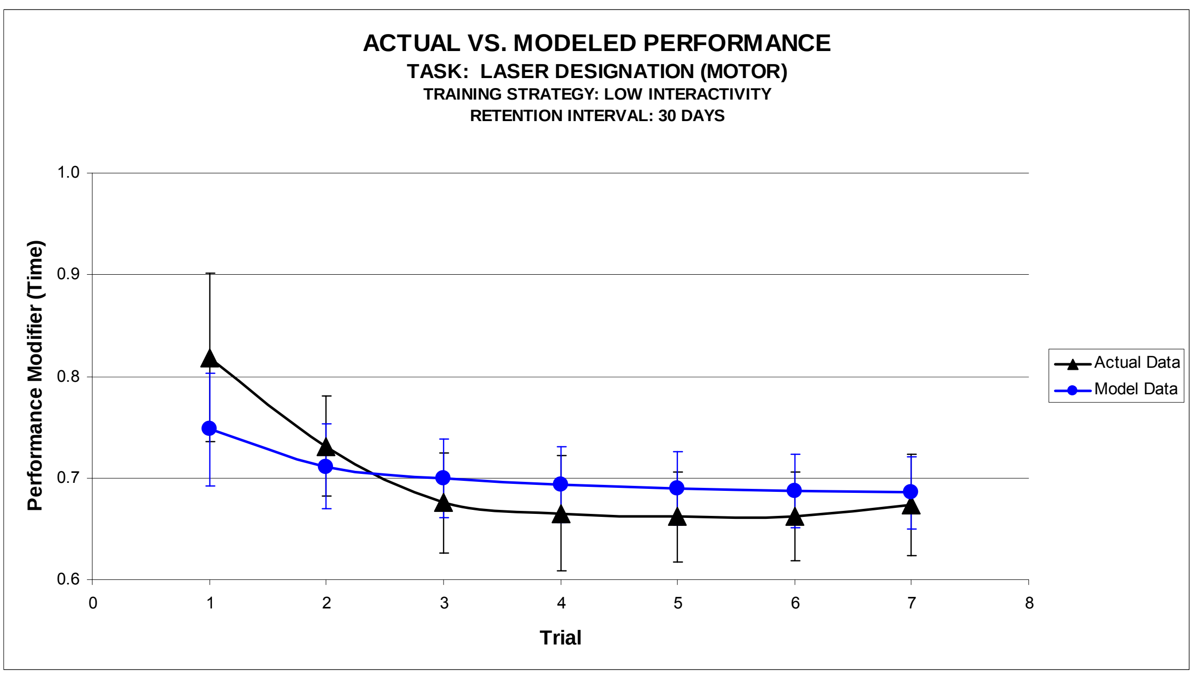


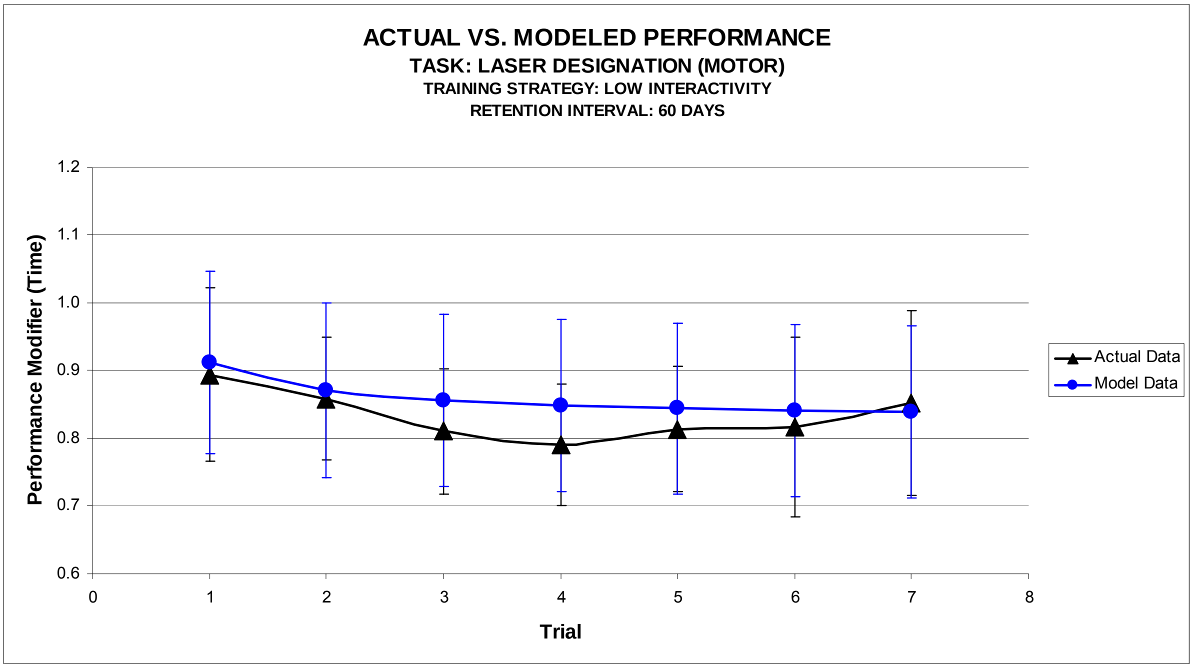
ACTUAL VS. MODELED PERFORMANCE
TASK: TOLD (COGNITIVE)
TRAINING STRATEGY: HIGH INTERACTIVITY
RETENTION INTERVAL: 30 DAYS
Performance Modifier (Time)
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

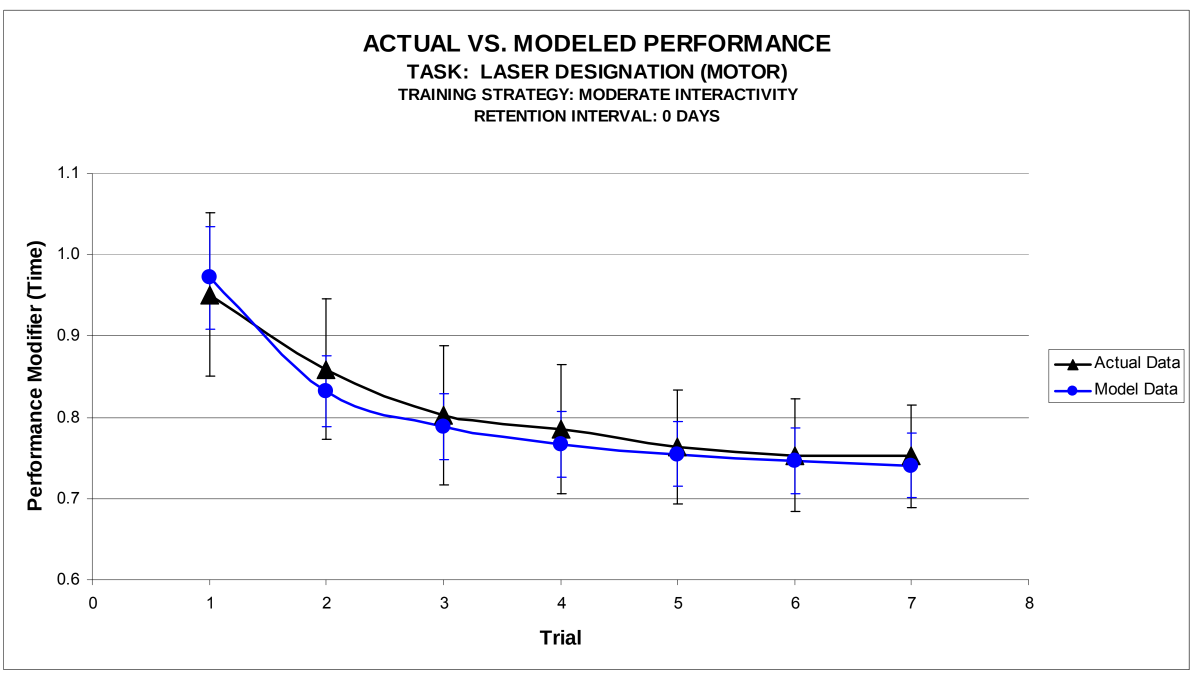

ACTUAL VS. MODELED PERFORMANCE
TASK: TOLD (COGNITIVE)
TRAINING STRATEGY: HIGH INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
2.5
2.0
1.5
1.0
0.5
0.0
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

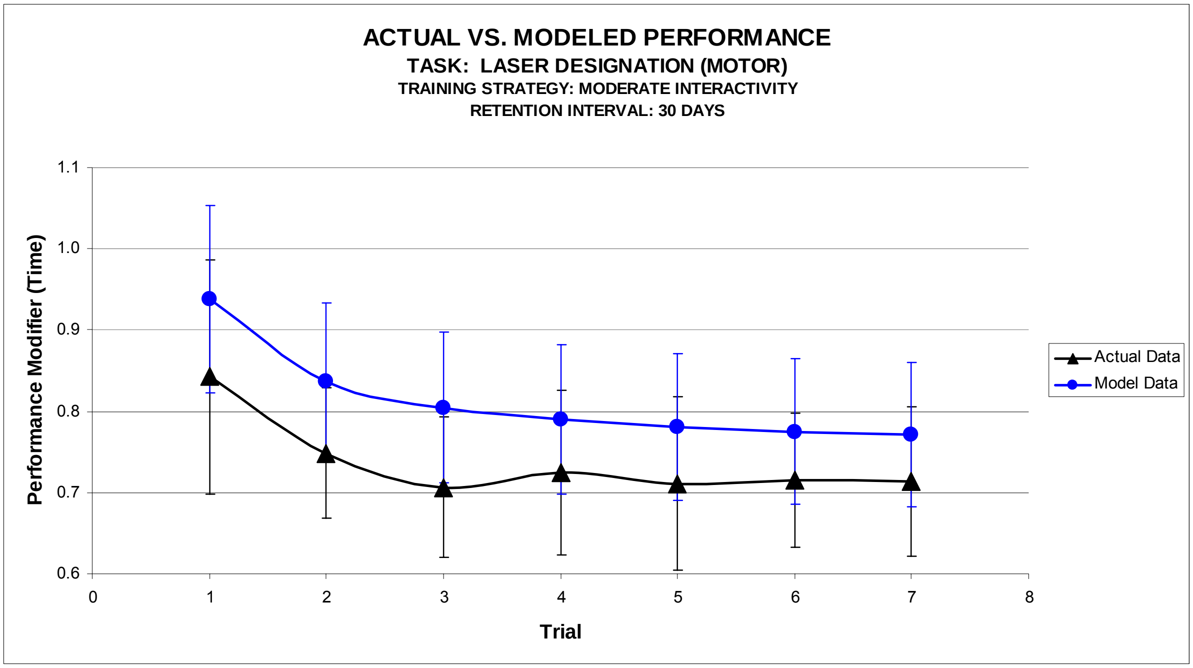
APPENDIX F: ACTUAL VS. FITTED DATA FOR MOTOR TASK (LASER DESIGNATION)

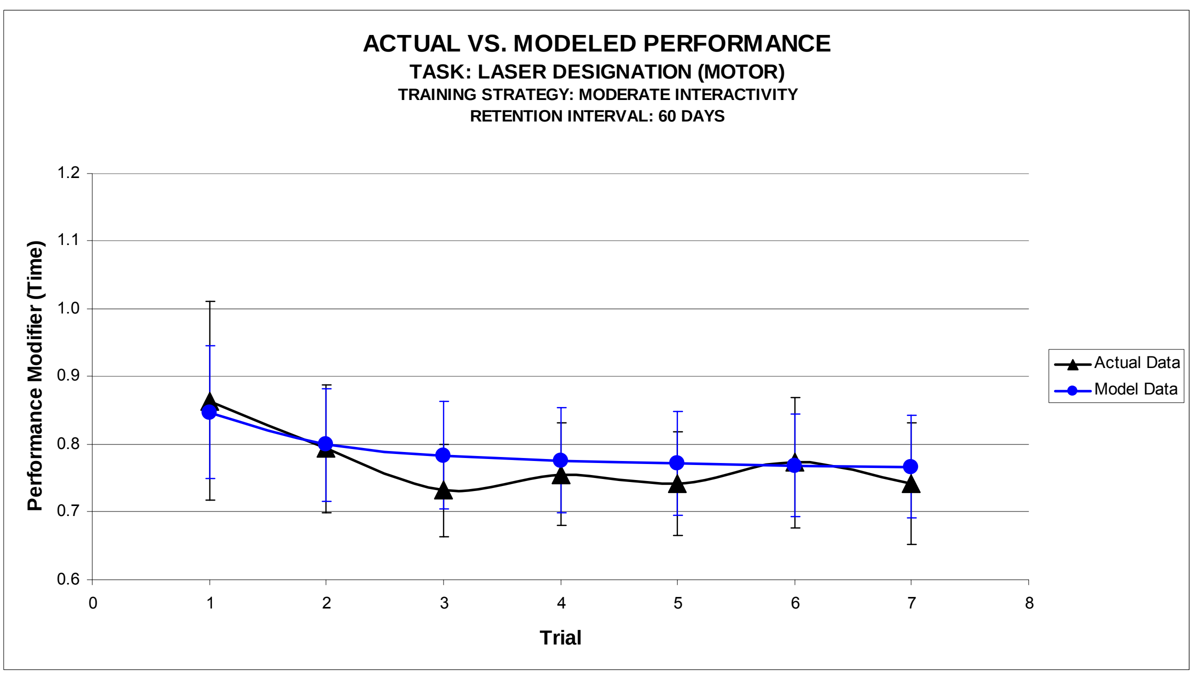


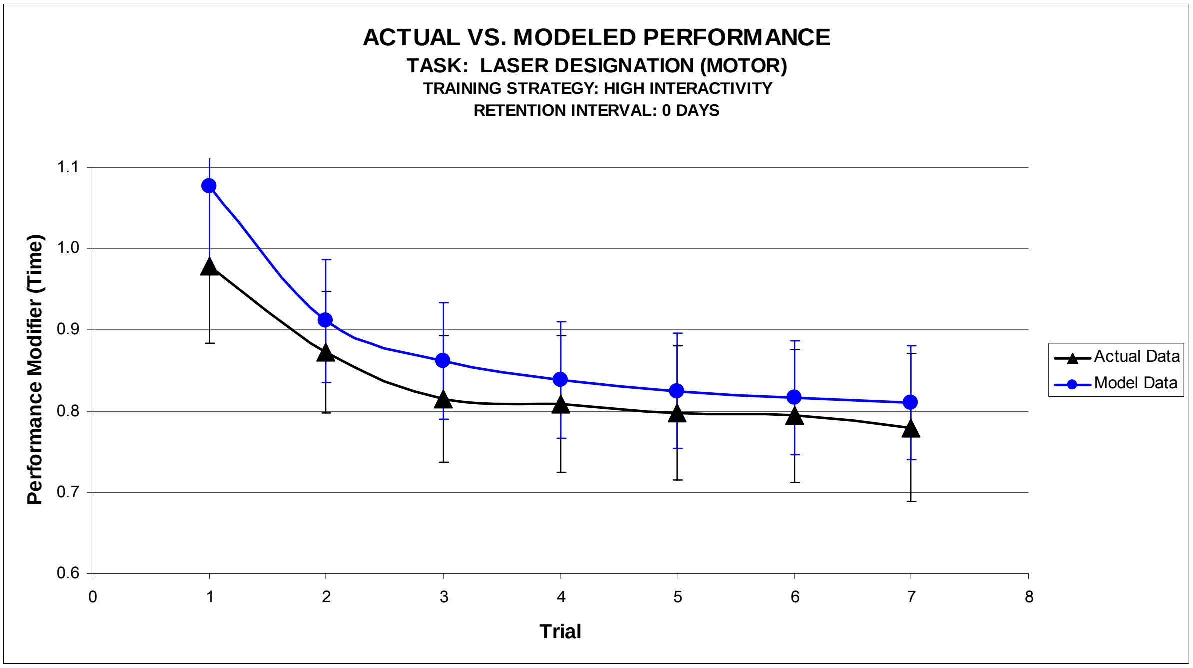


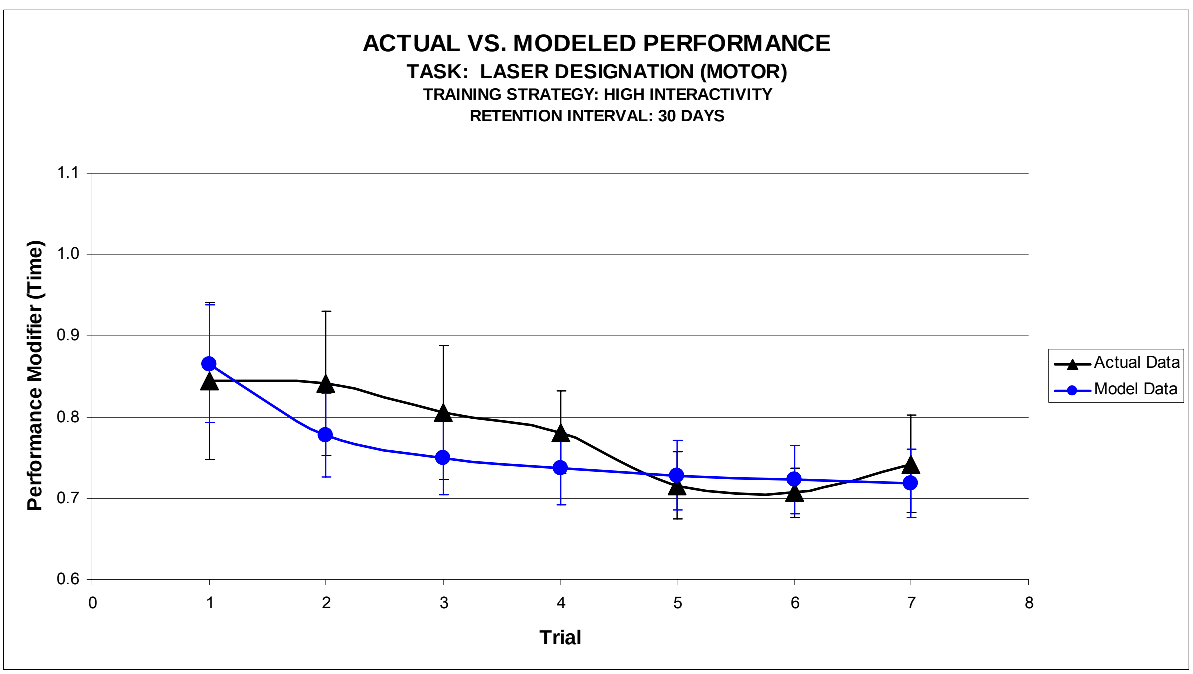

ACTUAL VS. MODELED PERFORMANCE
TASK: LASER DESIGNATION (MOTOR)
TRAINING STRATEGY: LOW INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
Trial
Actual Data
Model Data
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0
1
2
3
4
5
6
7
8

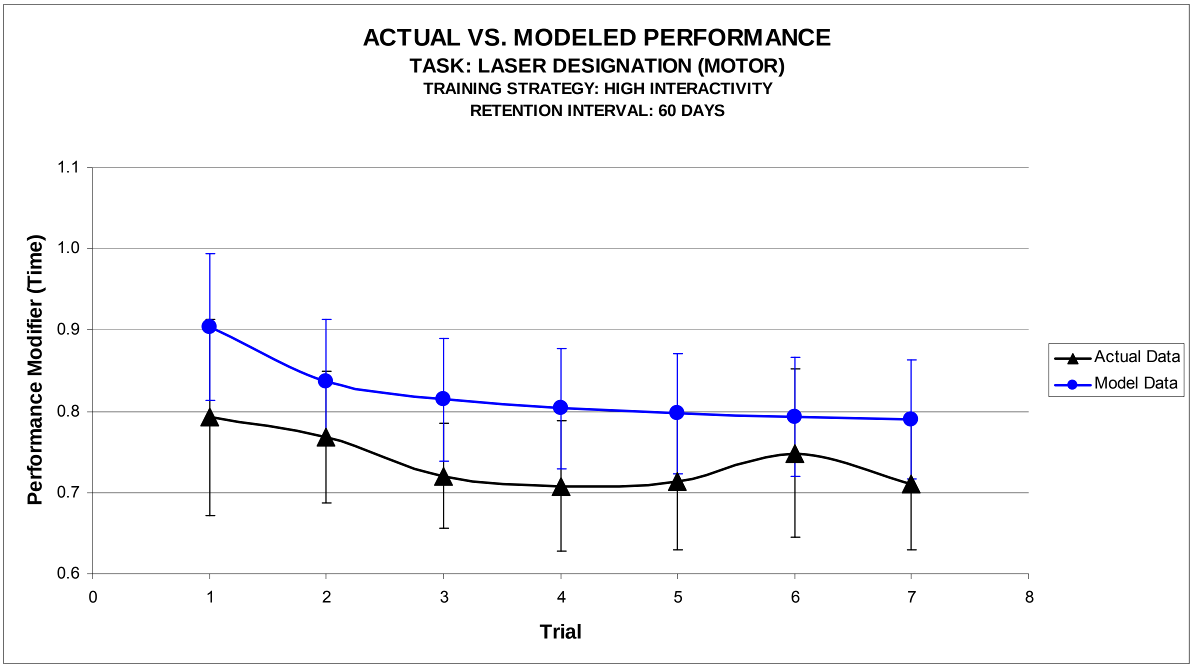

ACTUAL VS. MODELED PERFORMANCE
TASK: LASER DESIGNATION (MOTOR)
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 0 DAYS
Performance Modifier (Time)
Trial
Actual Data
Model Data
1.1
1.0
0.9
0.8
0.7
0.6
0
1
2
3
4
5
6
7
8


ACTUAL VS. MODELED PERFORMANCE
TASK: LASER DESIGNATION (MOTOR)
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 30 DAYS
Performance Modifier (Time)
1.1
1.0
0.9
0.8
0.7
0.6
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

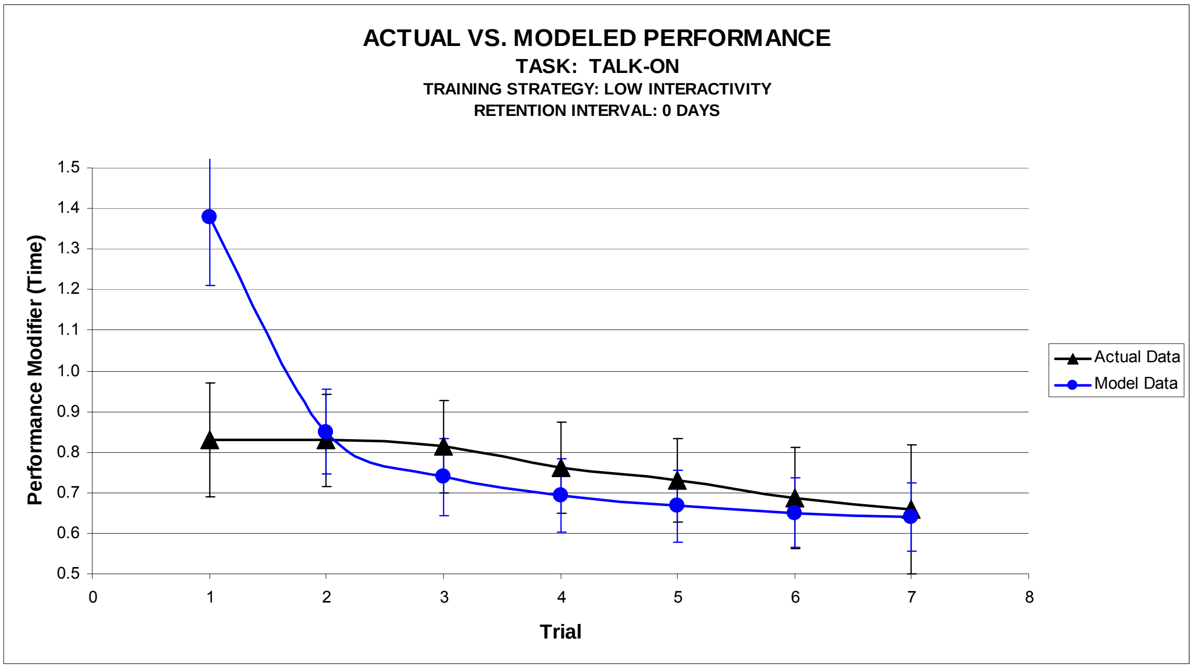

ACTUAL VS. MODELED PERFORMANCE
TASK: LASER DESIGNATION (MOTOR)
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

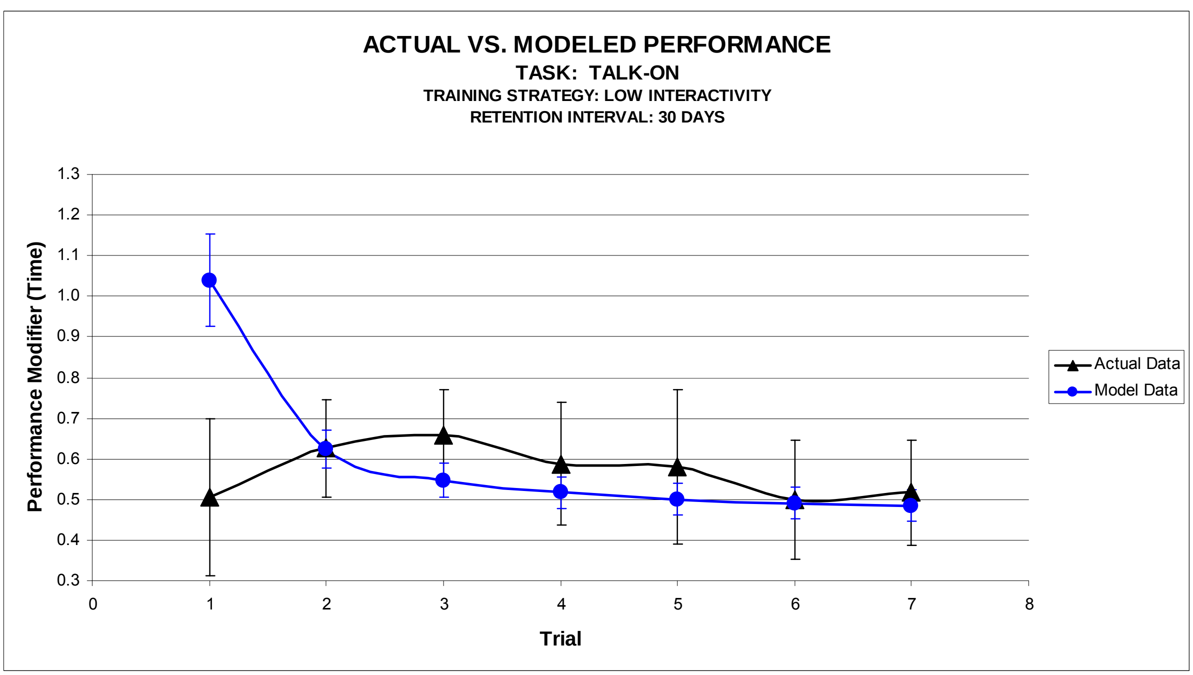

ACTUAL VS. MODELED PERFORMANCE
TASK: LASER DESIGNATION (MOTOR)
TRAINING STRATEGY: HIGH INTERACTIVITY
RETENTION INTERVAL: 0 DAYS
Performance Modifier (Time)
Trial
Actual Data
Model Data
1.1
1.0
0.9
0.8
0.7
0.6
0
1
2
3
4
5
6
7
8

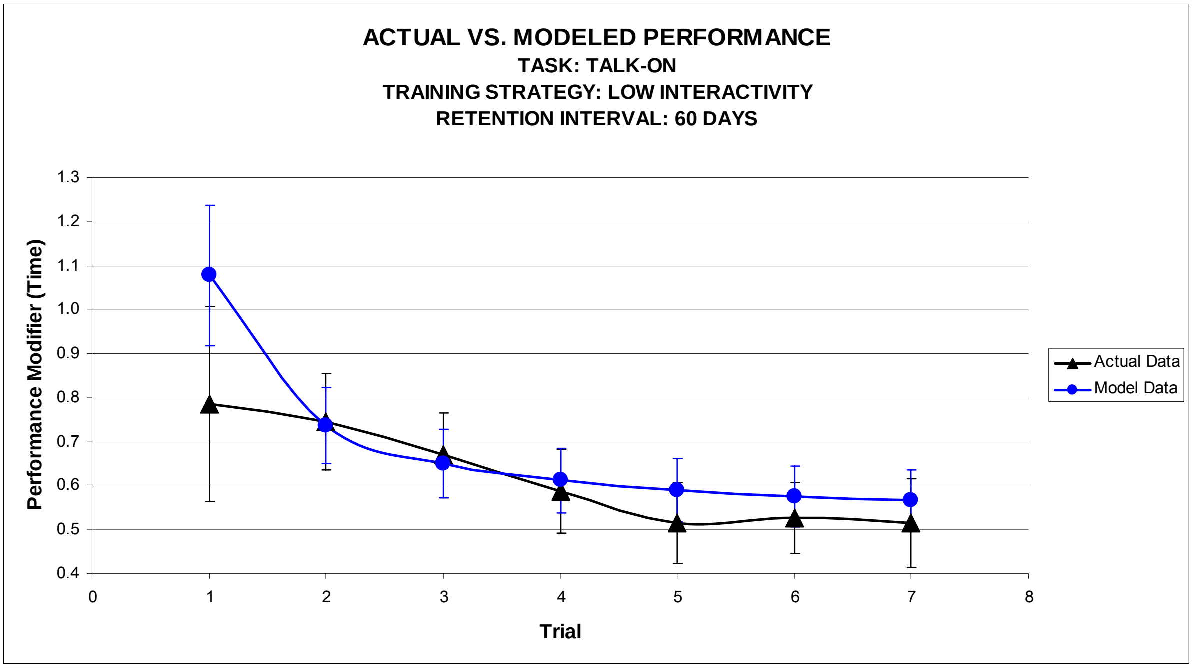

ACTUAL VS. MODELED PERFORMANCE
TASK: LASER DESIGNATION (MOTOR)
TRAINING STRATEGY: HIGH INTERACTIVITY
RETENTION INTERVAL: 30 DAYS
Performance Modifier (Time)
Trial
Actual Data
Model Data
1.1
1.0
0.9
0.8
0.7
0.6
0
1
2
3
4
5
6
7
8

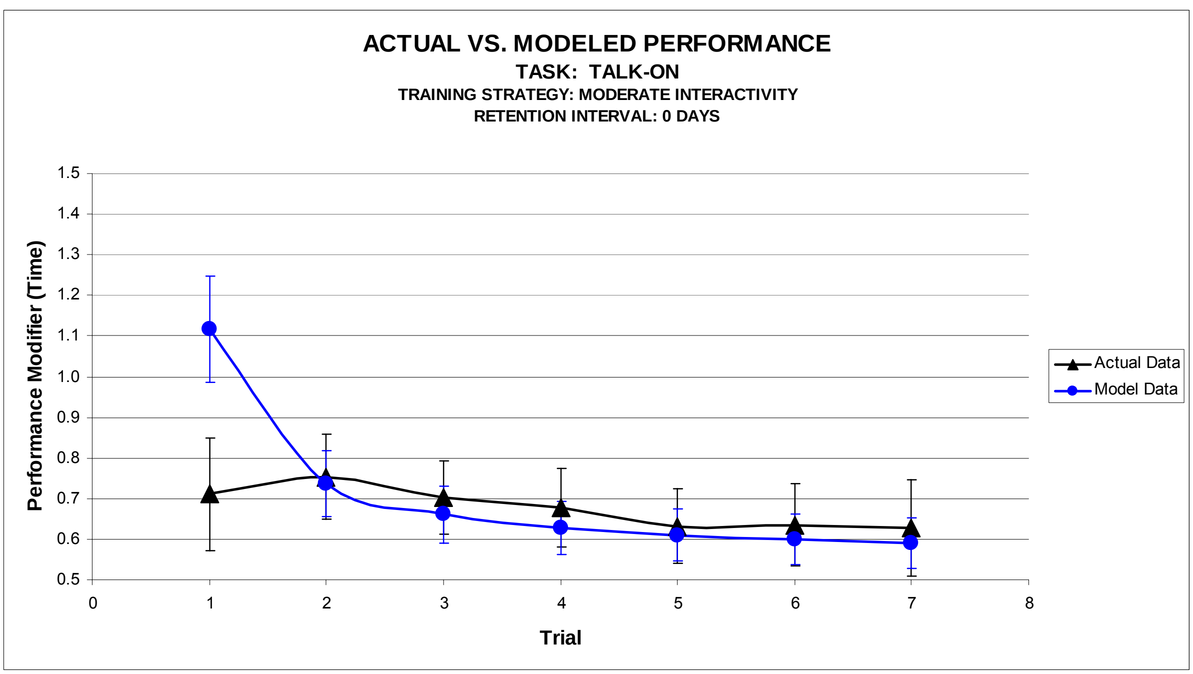

ACTUAL VS. MODELED PERFORMANCE
TASK: LASER DESIGNATION (MOTOR)
TRAINING STRATEGY: HIGH INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
1.1
1.0
0.9
0.8
0.7
0.6
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

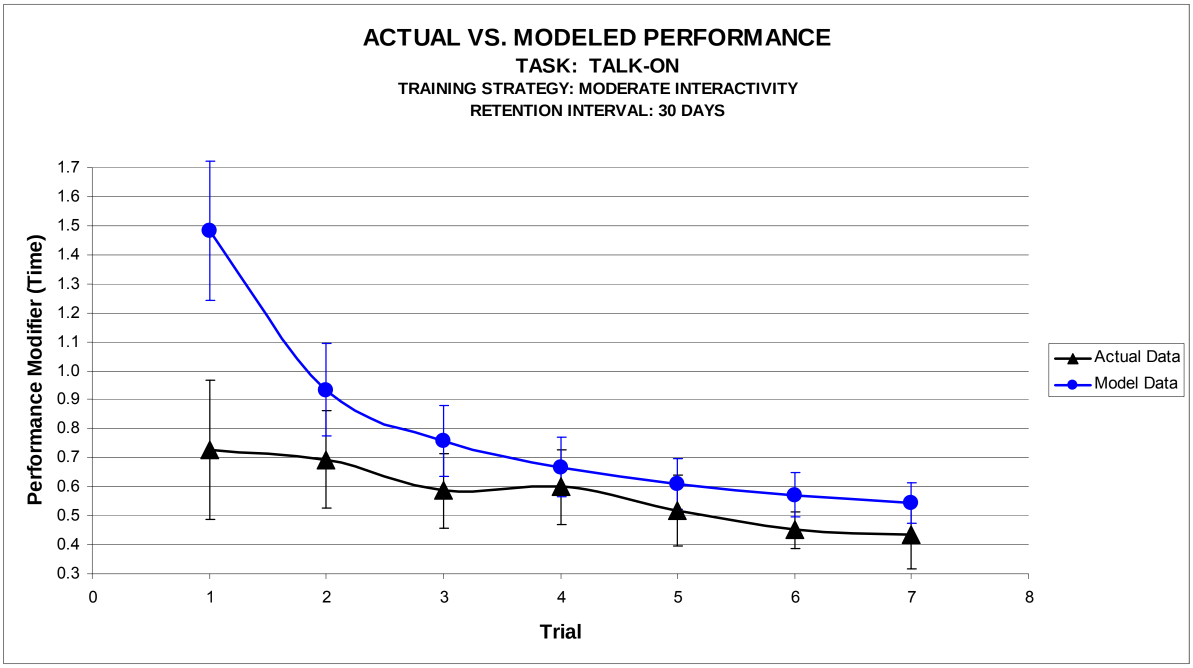
APPENDIX G: ACTUAL VS. FITTED DATA FOR PERCEPTION, COGNITIVE AND MOTOR TASK (TALK-ON)

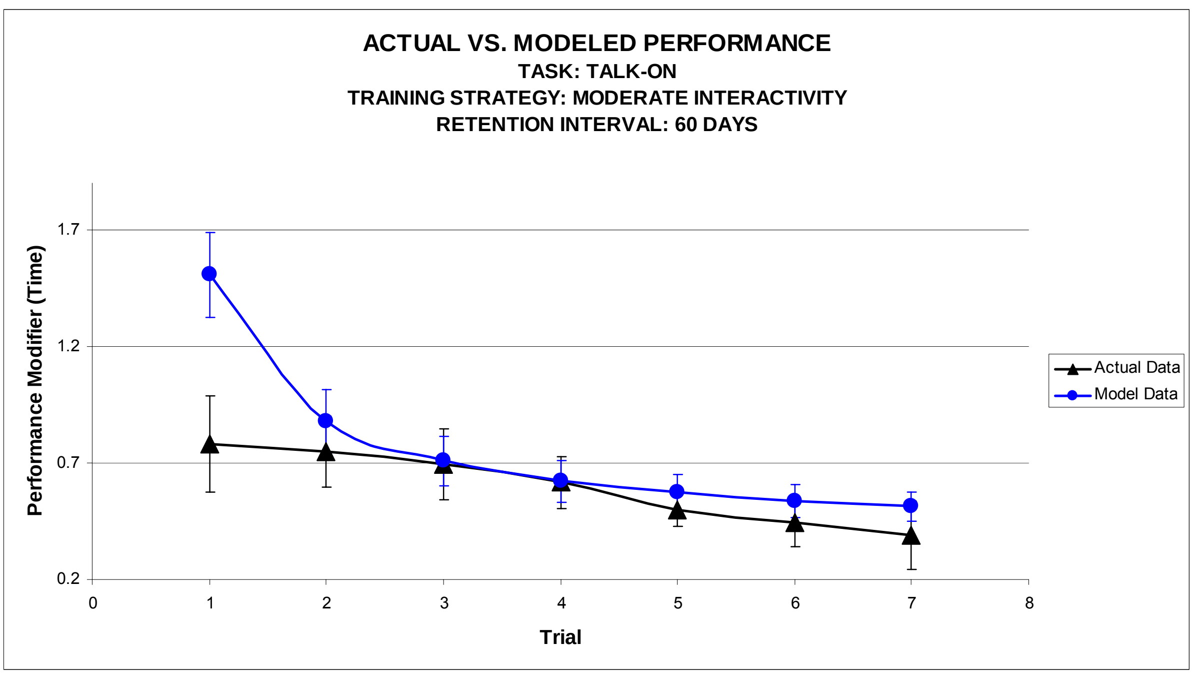


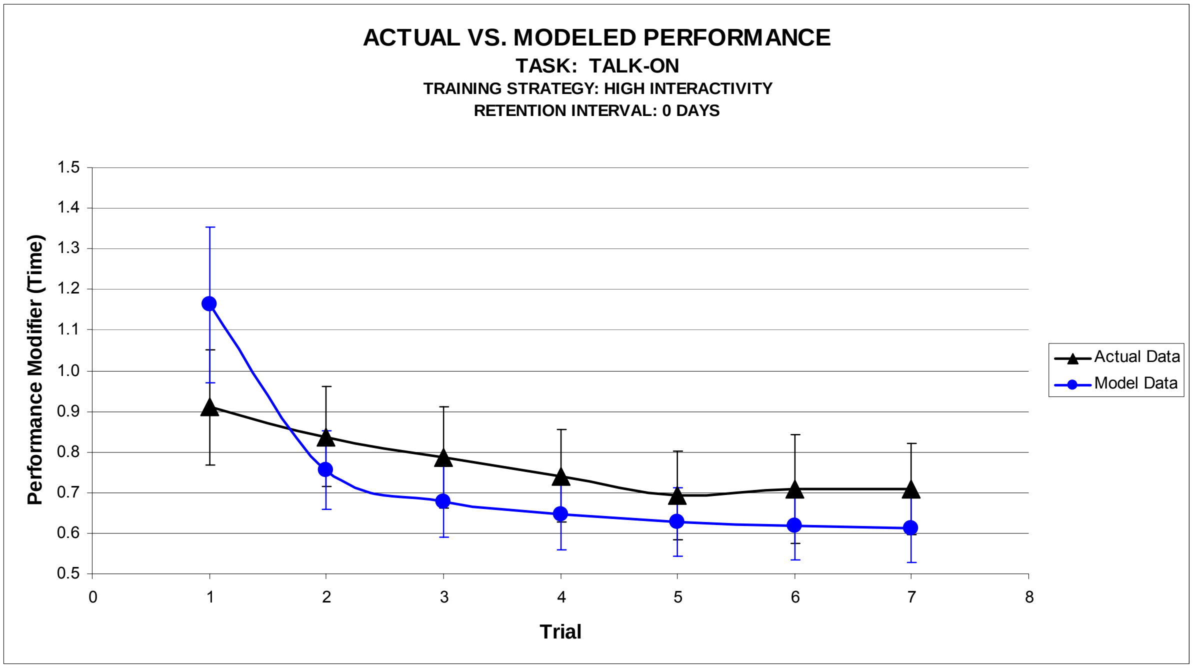


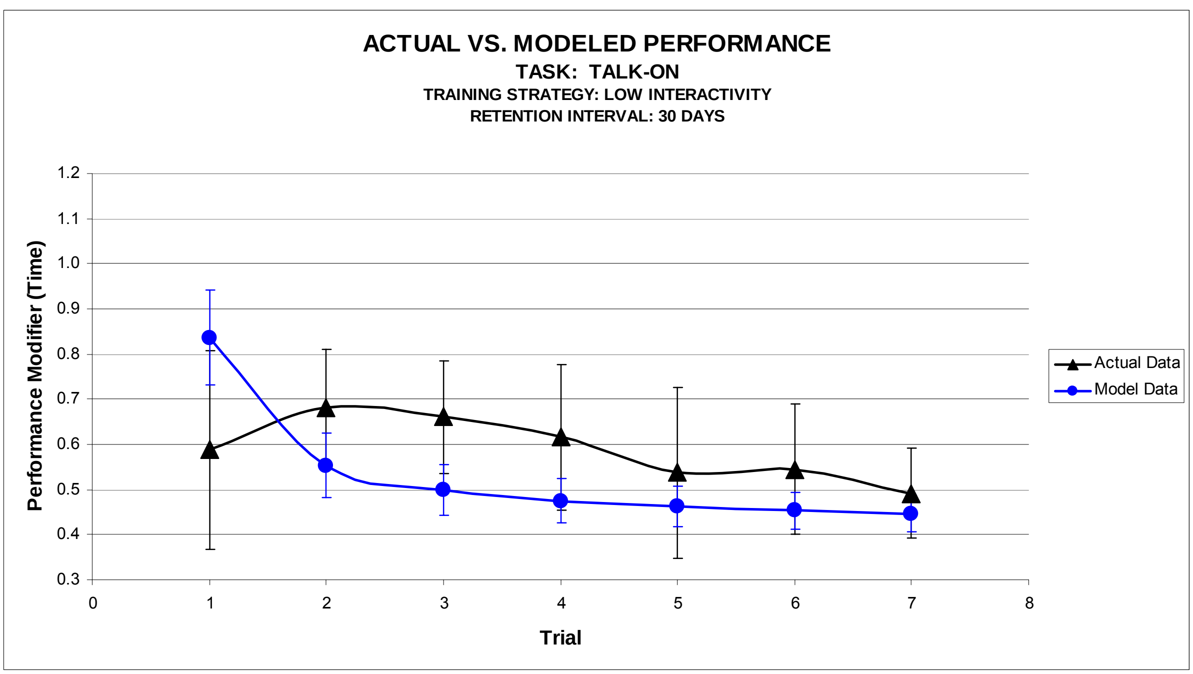

ACTUAL VS. MODELED PERFORMANCE
TASK: TALK-ON
TRAINING STRATEGY: LOW INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
Trial
Actual Data
Model Data

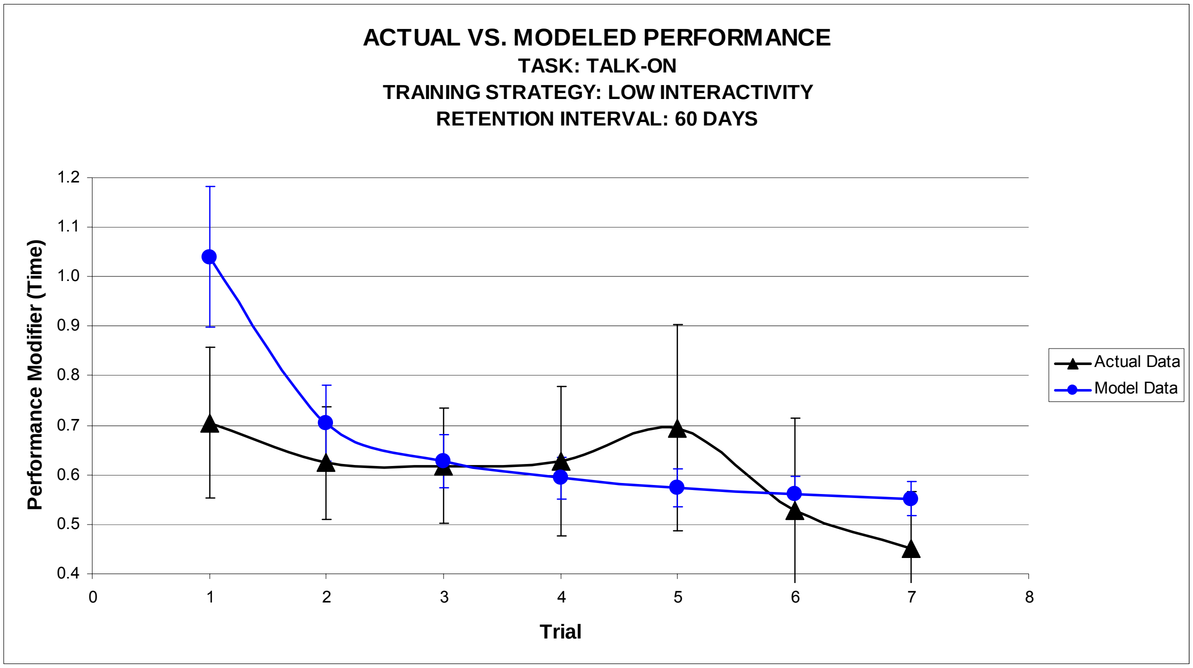

ACTUAL VS. MODELED PERFORMANCE
TASK: TALK-ON
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 0 DAYS
Performance Modifier (Time)
Trial
Actual Data
Model Data


ACTUAL VS. MODELED PERFORMANCE
TASK: TALK-ON
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 30 DAYS
Performance Modifier (Time)
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data


ACTUAL VS. MODELED PERFORMANCE
TASK: TALK-ON
TRAINING STRATEGY: MODERATE INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
1.7
1.2
0.7
0.2
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data


ACTUAL VS. MODELED PERFORMANCE
TASK: TALK-ON
TRAINING STRATEGY: HIGH INTERACTIVITY
RETENTION INTERVAL: 0 DAYS
Performance Modifier (Time)
Trial
Actual Data
Model Data


ACTUAL VS. MODELED PERFORMANCE
TASK: TALK-ON
TRAINING STRATEGY: LOW INTERACTIVITY
RETENTION INTERVAL: 30 DAYS
Performance Modifier (Time)
Trial
Actual Data
Model Data



ACTUAL VS. MODELED PERFORMANCE
TASK: TALK-ON
TRAINING STRATEGY: LOW INTERACTIVITY
RETENTION INTERVAL: 60 DAYS
Performance Modifier (Time)
1.2
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0
1
2
3
4
5
6
7
8
Trial
Actual Data
Model Data

REFERENCES

Ainsworth, J. S. (1979). *Symbol learning in navy technical training: An evaluation of strategies and mnemonics* (TAEG Report No. 66). Orlando, FL: U. S. Navy Training Analysis and Evaluation Group.

Allendar, L. (2000). Modeling human performance: Impacting system design, performance, and cost. In M. Chinni (Ed.) *Proceedings of the Military, Government, and Aerospace Simulation Symposium, 2000 Advanced Simulation Technologies Conference* (Washington, D.C.) 139-144.

Alion Science and Technology. (2007). *IMPRINT pro user guide volumes 1-3*. Boulder, CO.

American Institute of Aeronautics and Astronautics. (2001). *Guide to human performance measurements* (ANSI/AIAA G-035A-2000). Reston, VA.

Ames, L.J., & Sondhi, R. (1988). Technology's impact on training fidelity. In *Proceedings of the IEEE 1988 National Aerospace and Electronics Conference*, pp 970-976.

Anderson, J. R. (Ed.) (1981). *Cognitive skills and their acquisition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Archer, R., Archer, S., & Warwick, W. (2002). *Synergy between human systems integration and training to improve CGF realism.* 11th Conference on Computer-Generated Forces and Behavior Representation. Retrieved from **http://www.sisostds.org/index.php?tg=fileman&idx=list&id=2&gr=Y&path=CGF-BRIMS%2F11th+CGF-BR%2F11th+CGF-BR+Papers+and+Presentations**

Archer, R., Walters, B., Oster, A., & Van Voast, A. (2002). *Improving soldier factors in prediction models*. Army Research Institute Technical Report 1132.

Arthur, W., Bennett, W., Stanush, P., & McNelly, T (1998). Factors that influence skill decay and retention: A quantitative review and analysis. *Human Performance, 11(1)*, 57-101.

Aslaksen, E., & Belcher, R. (1992). *Systems engineering*. New York: Prentice Hall.

Bailey, R.W. (1989). *Human performance engineering: Using human factors/ergonomics to achieve computer system usability*. Englewood Cliffs, New Jersey: PTR Prentice Hall.

Chapanis, A. (1996). *Human factors in systems engineering*. New York: John Wiley & Sons.

Cousineau, D., Helie, S., & Lefebvre, C. (2003). Testing curvatures of learning functions on individual trial and block average data. *Behavior Research Methods, Instruments, & Computers, 35(4)*, 493-503.

Defense Acquisition University (DAU). (2006). *Defense acquisition guidebook.* Retrieved from **https://akss.dau.mil/dag/help_welcome.asp**

Department of Defense (DOD) (2003). DODI 5000.2 Operation of the Defense Acquisition System. Washington, D.C.: Government Printing Office.

Farr, M. J. (1986). *The long-term retention of knowledge and skills: A cognitive and instructional perspective* (AD-A175 905). Alexandria, VA: Institute for Defense Analyses.

Fitzgerald, P.C., Andrews, D.H., Crabtree, M.S., Doyal, J.A., Meador, D.P., & Walsh, W.J. (2008). *Incorporating training effects in modeling and simulation software*. In Proceedings CSER 2008, Conference on Systems Engineering Research, in press.

Heathcote, A., & Brown, S. (2000). The power law repealed: the case for an exponential law of practice. *Psychonomic Bulletin & Review, 7(2),* 185-207.

Lane, N.E. (1987). *Skill acquisition rates and patterns: Issue and training implications*. New York: Springer-Verlag.

Laughery, K.R., & Corker, K. (1997). Computer models and simulation. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (1375-1408). New York: John Wiley & Sons.

Law, A.M., & Kelton, W.D. (1991). *Simulation modeling and analysis*. New York: McGraw-Hill.

Lowry, J. C., Rappold, V. A., & Copenhaver, M. M. (1992). *Feasibility study for predicting human reliability growth through training and practice.* ARI Research Note 92-39.

Meador, D.P., Doyal, J.A., & Hale, C.R. (2008). *IMPRINT training algorithm enhancements (ITAE) for operators project interim report: Investigation of instructional strategy effects on operator performance*. Unpublished report submitted to AFRL/RHA.

Moroney, W.F., & Bittner, A.C. (1995). Military systems techniques. In J. Weimer (Ed.), *Research techniques in human engineering* (363-438). Englewood Cliffs, New Jersey: Prentice Hall.

Newell, A., & Card, S. (1985). The prospects for psychological science in human-computer interaction. *Human-Computer Interaction, 1(3)*, 209-242.

Newell, A., & Rosenbloom, P. (1981). Mechanisms of skill acquisition and the law of practice. In J. Anderson (Ed.), *Cognitive skills and their acquisition* (1-55). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

NIST, (2006). *NIST/SEMATECH e-handbook of statistical methods*. Retrieved April 4, 2008, from http://www.itl.nist.gov/div898/handbook/.

Orlansky, J., String, J., & Chatelier, P. (1985). *The cost effectiveness of military training*. Symposium on military value and cost effectiveness of training, pp. 97-109.

Rose, A.M., Czarnolewski, M.Y., Gragg, F.E., Austin, S.H., Ford, P., Doyle, J., & Hagman, J.D. (1985). *Acquisition and retention of soldiering skills*. ARI Technical Report 671.

Rouse, W.B. (1990). Training and aiding personnel in complex systems. In H. Booher (Ed.), *MANPRINT: An approach to systems integration* (417-432). New York: Van Nostrand Reinhold.

Salas, E., Burgess, K.A., & Cannon-Bowers, J. A. (1995). Training effectiveness techniques. In J. Weimer (Ed.), *Research techniques in human engineering* (439-475). Englewood Cliffs, New Jersey: Prentice Hall.

Sauer, J., Hockey, G.R.J., & Wastell, D.G. (2000). Effects of training on short- and long-term skill retention in a complex multiple-task environment. *Ergonomics, 43(12)*, 2043-2064.

Swezey, R.W., & Llaneras, R. F. (1997). Models in training and instruction. In G. Salvendy (Ed.), *Handbook of human factors and ergonomics* (514-577). New York: John Wiley & Sons.

Thurstone, L. (1919). The learning curve equation (Doctoral dissertation, University of Chicago, 1919).

United States Air Force (USAF) (1993). Air Force Manual (AFMAN) 36-2234 *Instructional System Development.*. Washington, D.C.: Government Printing Office.